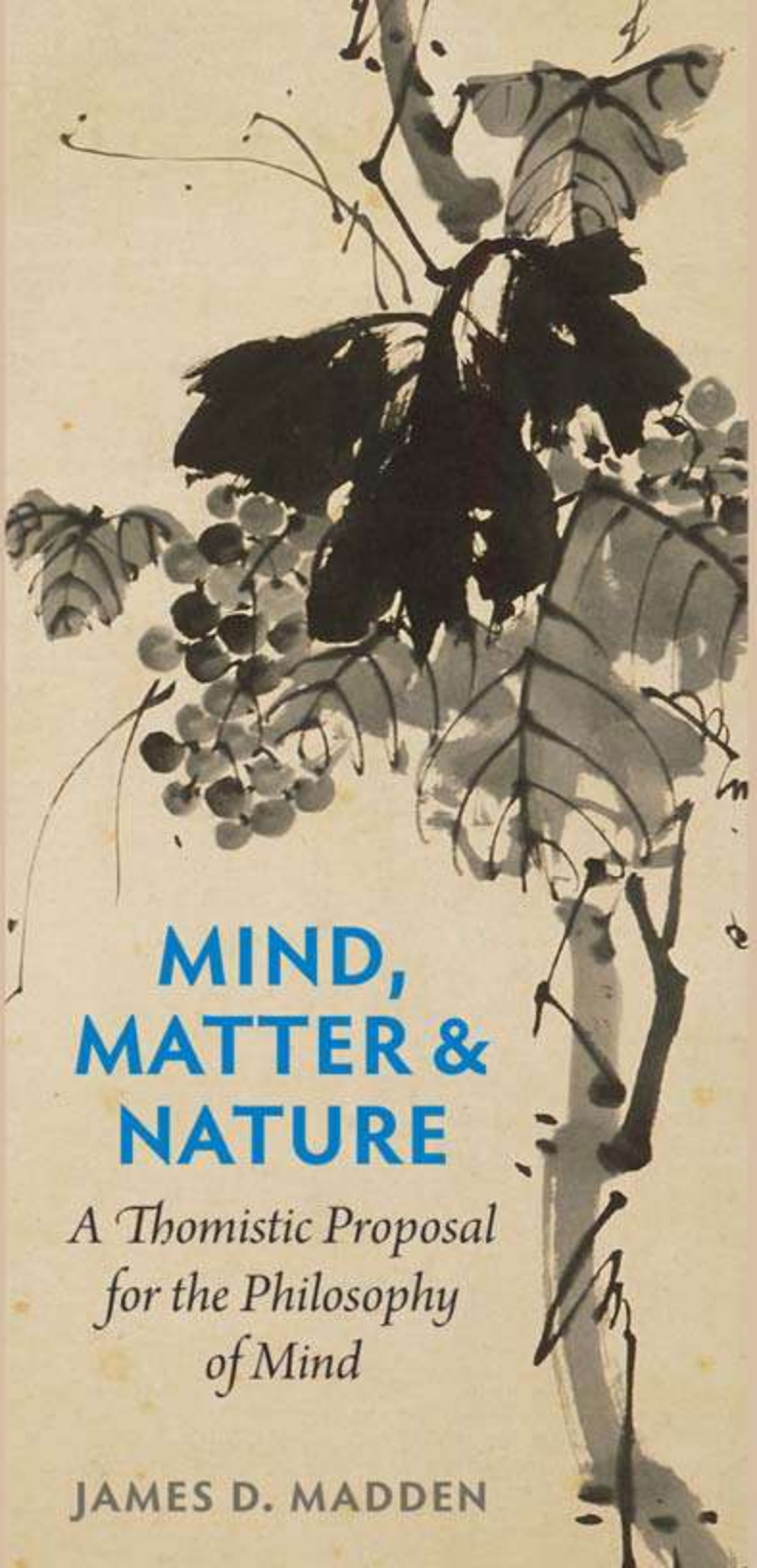

MIND, MATTER & NATURE

A Thomistic Proposal for the Philosophy of Mind

JAMES D. MADDEN

MIND, MATTER & NATURE

JAMES D. MADDEN

MIND, MATTER & NATURE

A Thomistic Proposal for the Philosophy of Mind

The Catholic University of America Press
Washington, D.C.

Copyright © 2013
The Catholic University of America Press
All rights reserved
The paper used in this publication meets the minimum
requirements of American National Standards for
Information Science—Permanence of Paper for
Printed Library Materials, ANSI Z39.48-1984.
∞

Library of Congress Cataloging-in-Publication Data
Madden, James D.
Mind, matter, and nature : a Thomistic proposal for the
philosophy of mind / James D. Madden.
pages cm
Includes bibliographical references (pages) and index.
ISBN 978-0-8132-2141-0 (pbk. : alk. paper) 1. Philosophy of
mind. 2. Naturalism. 3. Dualism. 4. Materialism.
5. Philosophy of nature. I. Title.
BD418.3.M3525 2013
128′.2—dc23 2012043720

For Jennifer

"Amor meus, pondus meum"

CONTENTS

Preface ix

CHAPTER 1
NATURALISM AND THE PHILOSOPHY OF MIND 1

What Is Naturalism? 1
Naturalism and the Philosophy of Mind 7

CHAPTER 2
THE CASE FOR DUALISM 19

What Is Dualism? 19
The Modal Argument for Substance Dualism 30
The Difference Argument for Dualism: Unity and Simplicity 47
Concluding Remarks about the Case for Dualism 57

CHAPTER 3
THE CASE AGAINST DUALISM: THE PROBLEM OF MIND-BODY INTERACTION 58

Intelligibility 60
The Law of the Conservation of Energy 65
The Causal Closure of the Physical 69
Psycho-Physical Dependence, Conscious Animals, and the Unity of Human Persons 81

CHAPTER 4
MATERIALISM 88

Eliminative Materialism 89
Behaviorism 95
Identity Theory 102
Token Identity Theory and Nonreductive Materialism 114
Functionalism and Artificial Intelligence 122

CHAPTER 5
PROBLEMS FOR MATERIALISM 132

Qualia Arguments—Bats, Cloistered Neuroscientists, and Zombies 133
Materialist Replies to the Qualia Arguments 138
Arguments from Thought—Beliefs, Intentionality, and the Man in the Chinese Room 147
Materialism and Agency 161

CHAPTER 6
EMERGENTISM AND NATURALISM 169

Emergence 171
Problems for Emergence 182
Emergentism and Abstract Thought 203
Emergentism and Agency 210

CHAPTER 7
BEFORE THE PHILOSOPHY OF MIND—THE PHILOSOPHY OF NATURE 217

The Problem of Change 222
Atomism and the Problem of Change 229
The Aristotelian Solution to the Problem of Change 235
Forms, Final Causes, and Scientific Explanation 243

CHAPTER 8
ARISTOTELIAN-HYLOMORPHIC PHILOSOPHY OF MIND 250

The Souls of Living Substances 251
The Souls of Conscious Substances 255
The Souls of Human Beings 265
But Is This Dualism? 274
A Key Objection and Reply 280
Some Tentative Conclusions 284

Bibliography 287
Index 305

PREFACE

ANOTHER OPINIONATED INTRODUCTION TO THE PHILOSOPHY OF MIND

We are not exactly suffering our way through a dearth of introductory books covering the philosophy of mind in the current academic and popular market. Many of these books are very helpful, leaving little room for improvement, and you will find that I cite some of them in this book. A few fairly common, though not universal, deficiencies in these books have moved me to write *Mind, Matter, and Nature: A Thomistic Proposal for the Philosophy of Mind.* Introductory texts in the philosophy of mind tend to take materialism, or at least some broad version of naturalism, as a default position. Even in those cases in which nonmaterialist philosophies are considered, this treatment rarely goes beyond a brief examination of Descartes's arguments in the *Meditations,* which are usually quickly dismissed by appeal to scientific progress or supposedly insurmountable problems of mind-brain interaction. The implication is often that we contemporaries just know better than to fall for something so silly as to believe in an immaterial soul. Such a hasty dismissal is unfair to Descartes and dualists in general, as there are arguments offered by contemporary philosophers, often inspired by Descartes, that go a long way toward supporting mind-body dualism.

Moreover, even when materialism is not simply presumed from the beginning, it is still common to accept the uniquely modern dichotomy between materialism and Cartesian dualism. The fact that there is a long history of philosophizing, that is, the Aristotelian-Thomistic tradition, that can be characterized as nei-

ther materialist nor dualist is frequently either blithely ignored or summarily dismissed as a prescientific bit of superstition. Introductions to the philosophy of mind by those sympathetic to the Aristotelian-Thomistic tradition are not always free from myopia of their own. Some Thomists tend to ignore the insights that contemporary analytic philosophy offers us in the philosophy of mind, and one also encounters some rather quick dismissals of both materialism and dualism. Moreover, when such books are conversant with contemporary philosophy of mind, they tend to try to fit Aristotelian-Thomism into the categories of such philosophizing. I believe that this maneuver, though well intentioned, has done much to motivate the charges of ambiguity and ad hoc gerrymandering that are often leveled against contemporary Thomists.

My intention in writing this book has been to avoid these pitfalls. I do not take philosophical naturalism as a default position, nor do I assume that our only options are Cartesian dualism and materialism. David Armstrong calls his book in which he gives an accessible defense of materialism *The Mind-Body Problem: An Opinionated Introduction,* and like him I present an opinionated introduction to the philosophy of mind, though from a very different perspective. Specifically, I argue that a broadly Aristotelian-Thomistic understanding of mind is our most viable option, even in light of the very best insights that contemporary analytic philosophy has to offer. This position, however, cannot be taken merely as a philosophy of mind, but rather as an application of a comprehensive metaphysics or philosophy of nature that endeavors to make sense of physical beings in general. Otherwise, I argue, Thomism does give every appearance of playing on obscurities.

Even though it is an *opinionated* introduction, this book is still an *introduction* to the philosophy of mind, and it is largely built out of materials I have used teaching undergraduate students. In chapter 1, I briefly introduce the reader to the philosophical nat-

uralism that dominates contemporary thinking about the mind and the problems it raises. In chapter 2 and chapter 3, I consider the arguments for and against dualism in some detail. In chapter 4 and chapter 5, I present the basic project that contemporary materialists set for themselves and the typical criticisms of their positions. In chapter 6, I consider emergentist versions of naturalism. Throughout these chapters I refer the reader to some of the most influential literature on the mind-body problem, and, when possible, I cite these articles as they appear in anthologies typically used in philosophy of mind courses. Thus, the reader of these chapters can expect to survey the basic lay of the land in contemporary philosophy of mind, and this material should fit in well with many of the sources typically read in undergraduate courses covering these issues.

In chapter 7, I argue that all of these various positions in the philosophy of mind seem to terminate in very similar conceptual dead ends, because they presume a common mechanistic philosophy of nature. I propose a version of Aristotelian hylomorphism as a philosophy of nature free of this assumption, which likewise solves fundamental philosophical problems. In chapter 8 I argue that hylomorphism, when understood as a philosophy of nature, averts the problems of both materialist and dualist philosophies of mind, because it eschews their common mechanistic assumption. I also answer some of the most common objections to the Thomistic understanding of mind by once again appealing to the broader context of the philosophy of nature in which it must be understood.

Though I am reluctant to use the phrase, what follows can be characterized as an exercise of what is now called "analytic Thomism." That is not to say that I believe that if St. Thomas Aquinas were alive today, he would be entirely pleased by all that I have to say in this book, nor do I claim that contemporary analytic philosophy can offer us insights into the interpretation of the Angelic Doctor's texts. I call this book an exercise

in analytic Thomism because I believe that the ultimate truth in the matters we are about to consider lies somewhere along the lines that St. Thomas outlines and that analytic methodologies are useful for making this fact apparent. My outlook on the relationship between these two traditions should not be taken as overly optimistic; though they are not utterly incommensurable, I do not believe that Thomism and analytic philosophy can be translated completely into each other's mother tongue. This is partly why I argue in the final two chapters of this book that to understand the Thomistic position, we must first ask and answer much broader questions that may be foreign to many contemporary philosophers.

I must acknowledge the debt of gratitude I owe to many people who have helped me in completing this project. W. Matthews Grant of the University of St. Thomas (St. Paul) has been generous with his time and encouragement when reading and offering helpful criticisms on various drafts of the manuscript. Whatever the value of this book, it is far better than it would otherwise have been without Professor Grant's kind assistance. I am also indebted to Charles Taliaferro and Patrick Toner, both of whom read the entire manuscript and offered helpful criticisms and suggestions. Comments from a group of anonymous reviews have been particularly helpful, though in some cases I am sure there will still be points of disagreement. Conversations with my colleagues at Benedictine College have been invaluable. Of particular note are Jean Rioux (from whom I have learned more about philosophy than I have from anybody else) and George Nicholas (a fine scholar of medieval literature, and the philosophy of mind's greatest autodidact), both of whom have endured many hours of my "thinking aloud." I have taught this material in a philosophical psychology course at Benedictine College for the last nine years, and a great many students have contributed to my thinking on these matters. Of particular note is Mackenzie Kilcawley, whose quick mind for logic always pushed me to

greater clarity when she was my student, and whose generosity with her time since graduating did me the great service of proofreading the entire manuscript. Andrew Jaeger, who has always been willing to talk the philosophy of mind with me, helped me a great deal, and it has been wonderful to watch him develop as a philosopher along the way. The intellectual integrity of Allison Thornton, John Bishop, and Elizabeth Carey has forced me to take my opponents' positions much more seriously than I otherwise would, and I believe that has borne fruits in this book. I all also appreciate the generous support I received from James C. Kruggel and the rest of the Catholic University of America Press editorial staff. Of course none of these people is responsible for this book's deficiencies, and please forgive me if there is somebody owed "thanks" whom I have forgotten.

Finally, I need to thank my wife, Jennifer, and our children, William, Martha, J. Patrick, Brendan, John, and Cormac. They have endured many summer days that I spent working on this book that might otherwise have been dedicated to more productive or just plain fun activities.

MIND, MATTER & NATURE

CHAPTER 1

NATURALISM AND THE PHILOSOPHY OF MIND

What Is Naturalism?

This book is both an introduction to the philosophy of mind and a critical reflection on the consequences of *naturalism* for our understanding of human nature. It is in fact quite difficult to separate recent philosophical discussions of the mind and the various versions of naturalism currently popular in the Western academy. "Given that naturalism is true, what sense can be made of the mind?" is really the question that most motivates many contemporary philosophers of mind. We do well then to begin our discussion by considering naturalism in some depth.

Naturalism enjoys widespread acceptance among contemporary Western intellectuals, even while the definition of "naturalism" suffers from a great many ambiguities.[1] The reason for this

1. For a detailed discussion of this controversy and its consequences for the coherence of naturalism, see Michael Rea, *A World without Design: The Ontological Consequences of Naturalism* (New York: Oxford University Press, 2002), 50–73.

state of unclarity is that "naturalism" is used to refer to a disparate collection of doctrines that have not necessarily been constructed as parts of a single systematic philosophy. For instance, classical atomist philosophers, such as Democritus and Epicurus, early modern empiricists such as David Hume, and more recent materialists all subscribe to something that may fairly be called "naturalism." Naturalism's popularity in Western universities coupled with the lack of consensus regarding its doctrinal essentials has led one of its prominent advocates in the early twentieth century to claim that even though "[w]e are naturalists now," this doctrine is nevertheless "of a very vague and general sort, capable of covering an immense diversity of opinion. It is an admission of a direction more than a clearly formulated doctrine."[2] I will not attempt what proponents of naturalism are not willing or able to do for themselves, namely, produce a rigorous definition that covers all naturalist doctrines without ambiguity. Instead, our purposes require us only to indicate the general direction of naturalist thinking, along with the key variants of naturalism that are relevant to the philosophy of mind in particular.

Intellectual fairness usually requires that one not take a definition of a philosophical doctrine from someone who is among its most outspoken critics. I will nevertheless ignore that otherwise sound rule of thumb for the moment. Even though C. S. Lewis is not a fan of naturalism (a great understatement!), his definition of naturalism is particularly helpful:

> What the naturalist believes is that the ultimate Fact, the thing you can't go behind, is a vast process in space and time which *is going on of its own accord.* Inside that total system every particular event … happens because some other event has happened; in the long run, because the Total Event is happening. Each particular thing … is what it is because other things are what they are; and so, eventually the whole system is what it is. All things and events are so completely interlocked that no one of them can claim the slightest independence from "the whole show." None of them exists

2. Roy Wood Sellars, *Evolutionary Naturalism* (Chicago: Open Court Press, 1922), vii.

"on its own" or "goes on of its own accord" except in the sense that it exhibits, at some particular place and time, the general "existence on its own" or "behavior of its own accord" which belongs to "Nature" (the great total interlocked event) as a whole.[3]

According to Lewis, the naturalist's basic claim is that nature *is going on of its own accord;* nature is *explanatorily self-sufficient;* that is, natural entities and processes provide us with ample resources to explain all that can be explained. When asked what we mean by nature, certainly no small question, Lewis's naturalist would reply that nature is the vast set of entities, events, and processes that occur within space and time, that is, those entities, processes and laws that we can discover using the scientific method, for example, physics, chemistry, biology, and the like. We may then more concisely say that in general *the naturalist believes that all events and entities within nature and nature itself are in principle explicable physically* (*where by "explicable physically" I mean "subject to the sort of explanation we give in natural science"*), *even if we do not have all such explanations given our current state of scientific development.* For example, if one accepts naturalism, then she would presume that the growth of plants, the motions of the planets, the development of political institutions, along with the very origin of the universe are either in principle physically explicable or not the sorts of things for which there can be significant explanations, at least for beings with our limited endowment of intelligence. Anything that supposedly falls beyond the purview of physical explanation is either nonexistent or something for which we can offer no rational explanation.

Although Lewis is a trenchant critic of naturalism, many contemporary naturalists think of their doctrine in very similar terms:[4]

3. C. S. Lewis, *Miracles* (New York: Harper Collins, 1947), 7–8.

4. Rea cites and discusses all three of these characterizations of naturalism, among many others, in *A World without Design,* 55–56, and argues that such characterizations of naturalism are wrought with difficulty. I agree with Rea, but this worry is beyond our current concerns. I will simply grant naturalists their way of characterizing their doctrine,

> Naturalism [is] the doctrine that reality consists of nothing but a single, all-embracing spatio-temporal system.[5]

> Naturalism [is] the belief that the world is a single system of things or events every one of which is bound to every other in a network of relations and laws, and … outside this 'natural order' there is nothing.[6]

> On the naturalist view, the world contains nothing supernatural … at the bottom level there are microphysical phenomena governed by the laws of microphysics, and, at higher levels, phenomena that not only participate in causal interactions describable in scientific laws but also bear the same general ontic relationship to microphysical items as do the entities quantified over and referred to [in] such higher-level laws as those which obtain in, for example, geology and neurophysiology.[7]

In all three renderings, naturalism is the view that everything that happens within nature can in principle be explained in the same categories that we use in the natural sciences, ultimately physics. We may then fairly claim that the general direction of naturalism is the claim that all events in nature, and nature itself, can in principle be given a physical explanation, inasmuch as they can be explained at all.

Scientism is the view that the scientific method is the only avenue to knowledge or justified belief.[8] Scientism certainly implies that all of reality (at least that which can be rationally understood) can be given a physical explanation. It is then not surprising that naturalism is at times defined in terms of scientism, that is, the denial of "the view that there exists or could exist any entities or events which lie, in principle, beyond the scope of sci-

---

whatever problems there may be in doing so. See Stewart Goetz, "Naturally Understanding Naturalism," *Faith and Philosophy* 27, no. 1 (2004):79–90.

5. David Armstrong, "Naturalism, Materialism, and First Philosophy," *Philosophia* 8, no. 2–3 (1978): 261.

6. Walter Stace, "Naturalism and Religion," *Proceedings and Addresses of the American Philosophical Association* 23 (1949–50): 22.

7. Michael Tye, "Naturalism and the Problem of Intentionality," *Midwest Studies in Philosophy* 19 (September 1994): 129.

8. One should not conflate scientism with a healthy respect for the explanatory power of modern science. One can consistently believe both that scientism is false and that the scientific method is a valid way of knowing. In denying scientism, one denies only that the scientific method is the *only* means for rational justification.

entific explanation."[9] Scientism and naturalism are then two ways of characterizing the view that "everything is a collection of entities of the sort the sciences are about, and all truth is determined ultimately by the truths about these basic scientific entities,"[10] or as Wilfred Sellars has famously put it, for the naturalist "science is the measure of all things, of what is that it is, and of what is not that it is not."[11]

There is a popular tendency to identify naturalism with what is traditionally called "materialism," and indeed there is some reason for doing so, since historically naturalists are usually materialists. There is indeed a close, though not necessary, connection between naturalism and materialism, but it will be crucial for subsequent discussions to realize that there are ways of being a naturalist without being a materialist in a strict sense. Materialism is a form of *ontological monism.* The prefix "onto" is derived from the Greek word for "being," so by saying that materialism is an ontological monism, I mean that it claims that there is only one fundamental kind of being, that is, material being. Although various versions of materialism have been held by philosophers throughout the millennia, all such views more or less amount to Epicurus's classical statement of the doctrine: the "totality is made up of bodies and void."[12] The most straightforward way of being a naturalist would be to be a materialist, especially because materialism would allow the naturalist to avoid sticky metaphysical questions regarding how nonphysical events or entities could be explained by physical entities or processes (much more on that worry later).

9. Arthur Danto, "Naturalism," in *The Encyclopedia of Philosophy,* ed. Paul Edwards (New York: Macmillan, 1967), 448.

10. John Post, *Metaphysics: A Contemporary Introduction* (New York: Paragon, 1991), 11.

11. Wilfrid Sellars, "Empiricism and the Philosophy of Mind," in *Science, Perception, and Reality* (Atascadero, Calif.: Ridgeview Publishing, 1991), 173.

12. Epicurus, *The Epicurus Reader,* ed. and trans. Brad Inwood and L. P. Gerson (Indianapolis: Hackett Publishing Co., 1994), 6. Benjamin Wiker provides a helpful, even if at times he is unfair to his opponents, history of materialism beginning with Epicurus in *Moral Darwinism: How We Became Moral Hedonists* (Downers Grove, Ill.: InterVarsity Press, 2002).

There have been, however, influential naturalists who explicitly disavow materialism. For example, Bertrand Russell, one of the twentieth century's best-known naturalists, presents a version of naturalism that is clearly not a version of materialism.[13] Naturalism, on the other hand, is typically taken as first and foremost a form of *explanatory monism.* Whereas materialism claims that there is fundamentally only one kind of being, that is, matter, naturalism claims that there is fundamentally *one kind of explanation,* that is, physical explanation, however many kinds of being there are. Along these lines, some naturalists have actually believed minds or souls to be immaterial, though ultimately accounted for physically.[14] The distinction between materialism and naturalism will become more important as we consider the various options available to the naturalist as she attempts to bring the mind into the framework of physical explanation. Since life for the naturalist would be easier if materialism were true, we will assume that materialism is the preferred version of naturalism, unless overwhelming difficulties force us to look for a more complicated variation.

Though naturalism may make no metaphysical claim directly, it certainly has implications for what we might reasonably believe to exist. Arguments for the existence of supernatural beings, for example, God or an immortal human soul, typically proceed a posteriori from a certain empirically given fact, for example, the existence of contingent beings or irreducible psychological phenomena, to a cause or explanation that transcends the explanatory domain of the laws of nature, Lewis's "Total Event." If there are no events that call for an explanation from outside the system of physical causes, then there would seem to be very little in the way of reasons for believing that there are any supernatural enti-

13. See Bertrand Russell, *What I Believe* (New York: Routledge, 2004), 1–8.

14. For a naturalist though nonmaterialist theory of the mind, see part 1 of Karl Popper and John Eccles, *The Self and Its Brain: An Argument for Interactionism* (New York: Routledge, 1977).

ties. Naturalism may not take a direct metaphysical stand, but it serves to undercut the typical routes for justifying belief in the supernatural. Thus, despite the pretension to metaphysical indifference espoused by many of its adherents, naturalism practically amounts to "the view that there is no such person as God, nor anyone or anything at all like him."[15] We might then finally define naturalism as the claim that *everything in nature that can be explained can be given a physical* (*or scientific*) *explanation, and the events, entities, and processes that constitute nature are all that we can reasonably believe to exist.*[16]

Naturalism and the Philosophy of Mind

Unless you have had some prior experience with academic philosophy, any understanding of naturalism you brought to reading this book likely came from the public controversy regarding the intelligent design movement. Proponents of intelligent design argue that naturalism fails because there are irreducibly complex biological systems that in principle cannot be explained in terms of evolution by natural selection. If there are such irreducibly complex biological systems (and not just those appearing to be irreducibly complex given our current state of scientific understanding), then it would seem that naturalism is in a great deal of trouble, because evolution by natural selection is the currently most plausible explanation of biological complexity consistent with naturalism. Other philosophers object to naturalism based on what they take to be the necessity of a first cause of the universe itself; they argue that nature is not self-sufficient, and therefore that it ultimately depends on some supernatural agency. Arguments against naturalism from irreducible complexity and the

15. Alvin Plantinga, *Warranted Christian Belief* (New York: Oxford University Press, 2000), 227.

16. Useful, though critical, introductions to naturalism can also be found in Stewart Goetz and Charles Taliaferro, *Naturalism* (Grand Rapids, Mich.: Eerdmans, 2008), and Charles Taliaferro and Jil Evans, *The Image in Mind: Theism, Naturalism, and the Imagination* (London: Continuum Publishing, 2011).

necessity of a first cause, and the defense against them by naturalists, provide immensely rewarding occasions for philosophical reflection, but they go beyond the more narrow concerns of this book.[17] Rather, our discussion will focus on the relation between naturalism and problems in the *philosophy of mind.*

Before attempting to define the philosophy of mind, it will be helpful to make a very basic distinction between *substance* and *property.* By "property" I mean the *way something exists,* a mode of being. For example, when we say "Will is tall," we claim that Will has the property of *being tall;* he exists in a tall way. When we say "The water is cold," we claim that the water has the property of *being cold;* it exists in a cold way. A substance, on the other hand, is something that is not a property of something else. A substance *has* properties, or properties *exist in* substances, but a substance (in this sense) is never had by, nor exists in, anything else. Substances are not ways of being, but entities that exist in certain ways. Will (who *is tall*) and the water (that *is cold*) are substances. Philosophers argue about which things are substances, what ultimate kinds of substances there are, and even whether there are any substances at all, but in this book we will concern ourselves only with a narrower set of concerns that arise in the philosophy of mind.[18]

17. A good introduction to first cause arguments can be found in William L. Rowe, *The Cosmological Argument* (New York: Fordham University Press, 1998). The best place to start on the argument against naturalism based on irreducible complexity is Michael Behe, *Darwin's Black Box* (New York: Free Press, 1998). I have written elsewhere regarding the limitations of the intelligent design argument against naturalism. See James D. Madden, "The Fifth Way, Scientism, and Intelligent Design," *Faith and Reason* 31, no. 3 (Fall 2006) 387–408; and James D. Madden and Mark A. Discher, "What Intelligent Design Does Not Imply" and "What Would Count as Defeating Naturalism? A Reply to Van Till," *Perspectives on Science and Christian Faith* (December 2004): 286–91, 296–98. See Daniel Dennett and Alvin Plantinga, *Science and Religion: Are They Compatible?* (New York: Oxford University Press, 2011) for a good overview of the debates surrounding naturalism, evolution, and intelligent design.

18. Useful introductions to the broader metaphysical issues to which I have alluded here can be found in Peter van Inwagen, *Metaphysics* (Boulder, Colo.: Westview Publishing, 2009), 23–108 , E. J. Lowe, *A Survey of Metaphysics* (Oxford: Oxford University Press, 2002), 23–79, and W. Norris Clarke, *The One and the Many: A Contemporary Thomistic Metaphysics* (Notre Dame, Ind.: University of Notre Dame Press, 2001), 94–161.

The philosophy of mind is concerned with accounting for *psychological beings* capable of sensations, thoughts, acts of will, reasoning, and the like. These distinctive characteristics of psychological beings are what we will call *psychological properties*. In our common practices, we intuitively distinguish between psychological properties and *physical properties* when giving explanations for our activities; that is, our ordinary ways of accounting for the behavior of psychological beings, including human beings, seem to imply that psychological properties and physical properties are fundamentally different modes of existence.[19] The propositions listed below refer to states of affairs of the sort we are readily acquainted with, but upon reflection we can see that they involve apparently different types of properties and explanations:

(1) Martha slipped because the sidewalk was icy.
(2) Martha was in pain after she fell.
(3) Martha winced after she fell because she was in pain.
(4) Martha thinks that the door is open.
(5) Martha thinks that the cat has escaped, because she thinks the door is open.
(6) Martha returned the candy, because she thinks honesty is good.

Proposition (1) doesn't seem, at least for common sense, to pose any deep problems. Martha's slipping is caused by the combination of her momentum while running and the low degree of friction between the soles of her shoes and the icy surface of the sidewalk. We are able to give a physical explanation for Martha's slipping in the same way we account for any nonhuman occurrence, that is, she slipped because of the arrangement of the relevant physical entities and the laws of nature. In fact, prior knowledge of Martha's momentum, the friction between the surfaces, and the relevant physical laws would allow us to predict the slip

19. This remark should not be taken as begging the question against forms of materialism that argue that psychological states or properties are physical states or properties. It might turn out that our common distinction is in fact inaccurate or needs to be reinterpreted. We will discuss these issues in great detail in chapter 4.

with a fair degree of accuracy. We would offer similar explanations in cases of Martha's tripping, sneezing, blinking, digesting, growing, and so on. Cases similar to (1) do not lead us to believe that human beings are in any way independent of the spatial-temporal system of physical entities and causes. The properties of Martha and the icy sidewalk that cause her slipping are all *physical properties,* where we mean by physical properties those characteristics of a substance that make some real difference to "at least part of the space which that thing occupies."[20] Due to her physical properties, Martha's slipping on the sidewalk can be accounted for as part of the Total Event just like any other event involving physical objects.

Consider (2), by which we claim that Martha has the property of *being in pain.* Martha's being in a pain is not quite so easily understood. Pain seems to be a different sort of thing from the low friction of an icy sidewalk or the momentum of a little girl running. Specifically, pain is a *psychological state* or *property* that doesn't seem to be straightforwardly something that can be located in space, nor is it something we can see or measure directly in the way we can do so for physical properties. The slipperiness of the icy sidewalk is readily available to anybody able to inspect the surface in question. More technically, the slipperiness of the sidewalk is apparent from a *third-person perspective.* Physical properties like slipperiness are *public* in the sense that there is no privileged perspective for observing them; the slipperiness is there to be inspected by anybody acquainted with the sidewalk. Martha's pain, on the other hand, is not available from a third-person perspective, but only from a *first-person perspective.* That is, only Martha can directly experience *her* pain. Martha's *being in pain* is *private,* not in the sense that her medical records are private (though not made publicly known, her medical records are physical objects that can be inspected from a third-person

20. E. J. Lowe, *Personal Agency: The Metaphysics of Mind and Action* (New York: Oxford University Press, 2008), 23.

perspective), but in a stronger sense such that her being in pain is inscrutable to everybody but Martha; only she can be directly aware of her pain. Of course we can have good reasons to believe that Martha is in pain, for example, her wincing, crying, cursing, and the like, but seeing evidence for pain is not the same thing as feeling the pain. This should be clear given (3), because we often take somebody's being in pain as a cause of his outward behavior. Wincing isn't pain, but rather the effect of pain.

Being in pain is not the only property that has this odd first-person status. For example, *Patrick sees red* or *Fido* (*a dog*) *smells dinner* attribute to Patrick and Fido properties (*seeing red* and *smelling dinner*) that are directly available only to Patrick and Fido. Privacy is an aspect of those psychological states we call *feelings* or *sensations*. Sensations just don't seem to be the same type of properties as slipperiness, momentum, mass, and such, nor do they fit very neatly in a universe that is supposed to be completely explained by the objective results of physical science. Of course there may be physical processes, for example, events in a brain, that are closely connected with sensations, but that is not to say that such processes and sensations are the same phenomena.[21] Moreover, as we can readily see in (3), sensations play a role in how we explain some human (and other animal) behaviors in ordinary situations. It makes perfect sense to say that Martha's *being in pain* caused her wincing, or that Fido's barking is caused by his smelling dinner. In short, we often explain behavior in terms of sensations.

In (4) we attribute a thought to Martha, and by "thought" I mean somebody's considering, or even asserting, a proposition or a certain state of affairs.[22] To say that *Martha thinks the door is open*

21. We will discuss this issue in great detail in chapters 4 and 5 below.

22. Some philosophers we will discuss use "belief" where I use "thought" in this book. I have avoided that usage, which is fine for the most part, because "belief" is often used by other philosophers exclusively to refer to an epistemic state that falls short of knowledge, e.g., "Brendan believes it will rain tomorrow" or "Will believes that Santa exists." Beliefs are a species of thoughts, but there are thoughts that are not beliefs, e.g., "Brendan thinks that 2 + 2 = 4." I realize that this choice of terminology on my part is not

is to say that she takes the proposition "The door is open" as true or the state of affairs of *the door's being open* as actually the case. Having a thought, like having a sensation, is private; nobody knows directly what Martha thinks except Martha. A thought is not the sort of thing we can inspect from a third-person perspective.

In addition to privacy, thoughts also exhibit *intentionality.* For Martha, "the door" *refers* to a particular door; that is, it is *about* some door. In this way we would say that the thought of a door has intentionality, because it points to or is about objects external to it. Martha's thought, like the word "door," *refers* to something beyond her, a door. Since utterances or inscriptions of words are ordinary material entities (ink marks on a page, vibrations of the air, images on a monitor screen, etc.), it would seem that intentionality doesn't raise any particular difficulties, but it is not as simple as that. Words have their intentionality only because we understand their meaning or interpret them. If there were no intelligent beings around to interpret a chance arrangement of physical objects in the shape of "door," these marks would not have intentionality. The intentionality of words is derived, whereas the intentionality of thoughts seems to be intrinsic to the person holding the belief. One may then wonder as to exactly how one physical object can be intrinsically (nonderivatively) about another physical object. The chair I'm sitting on isn't about anything, nor is the door about which Martha is thinking referring to something else. Intentionality does not appear to be a straightforwardly physical property, so thoughts don't seem to fit neatly into the Total Event. As John Haldane puts it, "the mental is intentionality: all mental phenomena have content, or are 'about' something or other."[23]

---

entirely satisfactory, because there are cases in which one thinks a proposition, e.g., I can think "Thomas Jefferson is the first President of the United States," without asserting it as true. Take it then that by a thought I mean any psychological state in which somebody entertains a proposition that could be asserted as true or false, including even those that are necessarily false or necessarily true.

23. John Haldane, "Analytic Philosophy and the Nature of Mind: Time for Another Rebirth," in *The Mind-Body Problem: A Guide to the Current Debate,* ed. Richard Warner and Tadeusz Szubuka (New York: Wiley-Blackwell, 1994), 199.

In virtue of their intentionality, thoughts can also be either true or false depending on whether the objects they are about exist in the ways they assert. If the door is indeed open, then Martha has a true thought, but if it is closed, her thought is false. It would be odd, for instance, to say in a straightforward sense that the chair I am sitting on or the door Martha is considering are true or false. Rather, our thoughts about these objects might be true or false. Being true is not something that seems to be a property of ordinary physical objects; we don't ordinarily speak of a *true door* or a *true chair* in the same sense in which thoughts might be true, though we may at times speak of a physical object being true in the sense of its being authentic ("That was a true lager beer") or reliable ("I can count on my tried and true fishing rod"). Some philosophers, for instance St. Thomas Aquinas, allow a sense in which physical objects are true, but only in terms of their "conformity with a human intellect"; that is, physical objects can be said to be true inasmuch as they cause us to have true thoughts about them.[24] Once again, being either true or false, properties that our thoughts have in virtue of their intentionality, does not seem to fit very easily into a universe understood entirely in terms of the resources of naturalism.

In this light, consider (5)–(6). In (5) we explain an inference on Martha's part in terms of a thought, whereas in (6) we explain an overt *action* on her part in a similar way. Let's call cases like (5) instances of *intellectual agency* and cases like (6) instances of *moral agency.* Just as thoughts seem to be distinct from ordinary physical states, actions explained by thoughts likewise don't easily fit into the rubric of physical explanation. This is obvious given the fact that we would hold Martha responsible for her performances in (5) and (6) in the sense that we would expect her to be

24. St. Thomas Aquinas, *On Truth,* trans. Robert W. Mulligan, SJ (Chicago: Henry Regnery Company, 1952), Q. I, A. 2. Saint Thomas also recognizes a sense in which physical objects can be true with respect to the divine intellect inasmuch as they conform to God's archetype or plan for them.

able to give *reasons* for her activities, but we likely would not do so for her behavior in (1). In most cases we would neither praise nor blame Martha for slipping on an icy sidewalk any more than we would do so for her sneezing. Slipping and sneezing are the same kind of event as a rock falling to the earth, and we certainly do not think to hold rocks responsible for their downward flights. Events like slipping, sneezing, and falling are explained by the antecedent condition of the physical universe and the relevant physical laws. Martha is not an agent in such cases; she doesn't do anything, but is merely acted on. When drawing inferences and making morally significant decisions, Martha is an agent; *she* does something. When drawing an inference, as in (5), it is her perceived truth of a set of propositions ("The door is open" and implicitly "Whenever the door is open, the cat has likely escaped") that leads her to infer another ("The cat has escaped"), and we would take such reasons as a fitting explanation in cases like (5). If there were no logical connection among these thoughts or one of her prior beliefs were obviously false, we would accuse Martha of making a poor inference, a gaffe of logic, or of being ignorant of something she ought to have known. If Martha's intellectual agency were just one more operation of the laws of nature through physical processes, this attitude of either praise or blame would be misplaced.

Likewise, we would applaud Martha's action in (6), because it is explained by her understanding of what is good and her decision to pursue such a good when she might have done otherwise. If this instance of moral agency were just one more case of physical causation, then applause for Martha's virtue would seem misplaced. We don't applaud physical objects for following physical laws. The fact that we think that we act, as least sometimes, in ways best explained through our own reasoning and decisions, and thereby we are responsible for such actions, is something quite foreign to physical explanation, and not what we would expect in a naturalist's universe. The point with respect to both (5)

and (6), and of course other similar cases, is that both intellectual and moral agency presuppose that our actions are not determined by prior physical causes, or at least this is how we typically think of ourselves as agents. That is, we seem to presume that neither our inferences nor our intentional actions are made to happen or are completely explained by prior states of the physical universe. Rather, such acts are explained, or justified, by our *thinking* that certain propositions are true or that certain actions serve the good. Once again, this apparent independence from prior physical explanation does not, on the face of the matter at least, seem like something explained as part of the Total Event.

Our foregoing discussion indicates that there are four facets of *human nature* that don't initially seem to square well with naturalism:

Human beings (presumably along with certain other animals) have sensations that explain some of their behaviors.

Human beings have thoughts that are either true or false.

Human beings are intellectual agents; some of our thoughts are explained not by prior physical causation, but by logical inferences from other thoughts.

Human beings are moral agents; some of our actions are explained not by prior physical causation, but by goods we choose to pursue.

Philosophy of mind is the attempt to arrive at the ultimate understanding of these facets of human nature. The philosopher of mind asks after what it really is to have sensations and thoughts and how it is that such psychological properties play a role in explaining our behaviors and actions (both logical inferences and acts of practical reason). The relevance of philosophy of mind to our assessment of naturalism should be readily apparent at this point. Throughout the foregoing discussion of sensation, thinking, intentionality, inference, and practical reason, I have pointed out

in each case that certain psychological states seem to fit poorly inside the spatio-temporal system of causes. I am sure that you have already sensed a tension between this discussion and the conception of naturalism we discussed in the previous section. It *seems* that sensations, thoughts, inferences, and practical reason defy the sort of explanation available to the naturalist. One might then conclude that these psychological properties are *nonphysical* or *mental,* and therefore stand outside the Total Event.[25] Note that I have not claimed, at this point anyway, that the naturalist cannot account for psychological states, only that on the face of the matter there are some difficulties that need to be addressed, if naturalism is to be a reasonable philosophical position: Can sensations and thoughts and their apparent explanatory power be accounted for in a naturalist worldview? If so, then human minds do not pose any serious evidence against naturalism. If not, then human minds would seem to pose devastating evidence against naturalism.

As I have mentioned earlier, we will take our default version of naturalism to be materialism in the early going of our discussion, though we will eventually discuss versions of naturalism that do not purport to be versions of materialism. Daniel Dennett, one of the most outspoken advocates of naturalism in the American academy, characterizes the agenda for the naturalist in the philosophy of mind as follows:

> There is only one sort of stuff, namely matter—the physical stuff of physics, chemistry, and physiology—and the mind is somehow nothing but a physical phenomenon. In short, the mind is the brain. According to the materialists we can (in principle!) account for every mental phenomenon using the same physical principles, laws and raw materials that suffice to explain radioactivity, continental drift, photosynthesis, reproduction, nutrition, and growth.[26]

25. In what follows, I will use the term "psychological" as ontologically neutral between physical and nonphysical, whereas I will use "mental" as having a nonphysical connotation.

26. Daniel Dennett, *Consciousness Explained* (Boston, Mass.: Back Bay Books, 1991), 33.

In other words, the naturalists, whom we will assume to be materialists for the time being, believe that psychological states, acts, and processes are material; they are in no way different from the material beings, states, acts, and processes that we find in physical systems such as crystals, plants, lower animals, and so forth.[27] Given that it seems that sensations and thoughts are nonphysical, a case for materialism often consists of, at least for starters, an attempt to explain how it is that *seemingly* nonphysical or mental things could turn out to be physical things. The success of materialism will bode well for naturalism, whereas its failure will count against the naturalist's case. Remember, however, that a defeat for materialism will not necessarily amount to a defeat for naturalism, if there are plausible nonmaterialist versions of naturalism that can be defended.

Whether human beings as sensing, thinking, and responsible animals can be fully integrated into a naturalist worldview is certainly a question the answer to which holds great import for our broad understanding of the meaning, purpose, and value of human life. We will not address these sorts of all-important questions in this book.[28] Rather, our inquiry is a necessary preamble to such questions; we will concern ourselves with the much more technical and even pedestrian concerns requisite to answering the broader "Who am I?" questions. In chapter 2, we will begin our inquiry by considering the case for substance dualism, that is, the view that human beings are composed of distinct material and nonmaterial substances, which exist and

27. There are materialists, so-called eliminative materialists, who would deny this statement, because they deny that there are any such things as psychological states. See chapter 4 for a discussion of this view.

28. For considerations of broad issues of meaning and value from a naturalist perspective, see Owen Flanagan, *The Problem of the Soul* (New York: Basic Books, 2002); B. F. Skinner, *Beyond Freedom and Dignity* (Indianapolis, Ind.: Hackett Publishing, 2002); and Richard Rorty, *Contingency, Irony, and Solidarity* (Cambridge: Cambridge University Press, 1989). C. S. Lewis famously criticizes these sorts of naturalistic accounts of meaning and value in *The Abolition of Man* (New York: Harper Collins, 2002), and a consideration of the aesthetic dimension of the case for and against naturalism can be found in Taliaferro and Evens, *The Image in Mind.*

operate independently. In chapter 3 and chapter 4 we investigate the arguments offered against dualism and the materialist alternatives that are proposed in its stead. We will then turn our attention in chapter 5 to criticisms of materialism, and in chapter 6 we will discuss nonmaterialist versions of naturalism, along with their requisite difficulties. We will begin our exploration of the Aristotelian-hylomorphist alternative to all of the positions we have discussed thus far in chapter 7, where we will consider the broader metaphysical considerations that lead to hylomorphism independently of the problems of the philosophy of mind. In chapter 8 we will conclude our discussion by investigating the implications of Aristotelian hylomorphism for the philosophy of mind in particular.

CHAPTER 2

THE CASE FOR DUALISM

What Is Dualism?

In the last chapter we defined naturalism as the doctrine claiming that everything can in principle be explained physically, and we concluded that materialism is the most straightforward application of naturalism to the philosophy of mind. The materialist claims that if there are such things as psychological substances, states, acts, processes, properties, and the like, then they are material things not ontologically different from other nonpsychological entities. What generally goes by the title of "dualism" is a denial of materialism. Specifically, the dualist argues that psychological substances, states, acts, processes, and the like, are nonphysical or mental things, ontologically different from nonpsychological material beings. In other words, while materialists are ontological *monists,* believing that there is only one kind of being, dualists maintain that there are two fundamentally different kinds of being, material and mental.

We may distinguish between two basic forms of dualism. The first is *property dualism:* the human mind is a collection of mental

properties had by the human body, specifically the brain. Mental properties, for example, sensations and thoughts, are ontologically different from physical properties, for example, mass, motion, charge, position, and for the property dualist, a conscious organism is a single physical substance that has both sets of these properties. Martha's pain sensation and her thinking *that the door is open* are distinct from her physical properties and play a role in explaining her behavior. Though we might find the notion of a material substance having nonphysical (mental) properties to be curious, property dualism does have the virtue of parsimony, because it only asks us to believe in one kind of substance.

A more robust version of dualism is *substance dualism:* human minds and bodies are distinct substances, the former being nonphysical and the latter being physical. According to the substance dualist, the mind is not just a collection of nonphysical properties, but an individual, nonphysical substance. The body is a physical substance just like any other tangible object, whereas the mind is a mental substance independent of any physical object for its existence and activity.[1] Sensations and thoughts, on this view, are nonphysical properties had by a nonphysical substance. Many substance dualists claim that even though the mind and the body are distinct substances, each occupying a distinct ontological domain (physical or mental), there can nevertheless be causal interaction between minds and bodies, such that minds can cause physical events in the body and the body can cause mental events in the mind. We will call this version of substance dualism *interactive dualism.*

Although dualists and materialists disagree on almost every substantive issue in the philosophy of mind, it is important for our discussions that these otherwise opposed philosophical

1. In chapter 6 we will discuss a position known as emergent dualism, according to some versions of which the mind is a distinct, nonphysical substance, but it is not entirely independent from the body. For the sake of simplicity, we will, for the moment, take substance dualism as including the claim that the mind is independent from the body.

points of view do often share an important point of agreement: *mechanism*.[2] Peter Unger has recently offered a characterization of mechanism in his description of the worldview, what he calls "Scientificalism," that underlies most modern philosophical reflection (including the philosophy of mind):

> First … there is physical stuff or *matter* … this matter is independent of minds: To exist the matter needn't be sensed by, or perceived by, any sentient beings. Second … Scientificalism says this: Insofar as it's determined by anything at all … the distribution of the world's matter at any given time is determined by, even if only probabilistically determined by, the distribution of the matter at earlier times, all of it proceeding in line with the world's basic natural laws, which are all physical laws. Third … some of the world's matter, or possibly much of the matter, is configured so as to compose various complex material structures and systems.… So, here we have stars, and planets, and rocks, and rivers. Among the more complex of even these highly complex material structures and systems are living entities, or those serving to constitute living entities. And, here we have plants and animals, including human animals.[3]

The mechanist (Unger's Scientificalist), takes the most basic or fundamental particles of matter that compose the natural world to be substantial in their own right; that is, they exist independently of anything else, especially minds (though he allows for the possible exception of an original creation by a divine mind).[4] Moreover, the whole story about physical objects, and subsequently the natural world, can in principle be told entirely in terms of physical properties, for example, position, velocity, mass, spin, of these fundamental particles of matter and the laws of nature. Nonquantifiable notions such as sensation and intentionality have no place in our understanding the natural world: physical particles, the pri-

2. In later chapters, we will encounter both dualists and materialists who either reject or greatly qualify mechanism. My only point here is to point out a crucial background assumption that has been quite common to much of modern philosophy of mind, and that is particularly relevant to understanding the development of the traditional arguments for substance dualism that I will discuss below.

3. Peter Unger, *All the Power in the World* (New York: Oxford University Press, 2006): 6–7.

4. Ibid., 6. Unger himself is not terribly sympathetic to the notion of creation; see Unger, *All the Power in the World,* 501–7.

mary furniture of the natural world, do not have such properties as thoughts and sensations, and we do not need to invoke these notions when it comes to explaining the behavior of such objects. In fact, the mechanist would prefer "to reduce all the qualities and activities of bodies to quantitative realities."[5] That is, mechanists typically claim that physical particles, and the complex physical substances that they compose, do not actually have any qualitative or generally nonquantifiable properties; as Unger puts it, the mechanist claims "there are not … any of the Qualities anywhere in physical reality."[6] Thus, for the mechanist, the physical world is exhausted by fundamental physical entities and complex aggregates of such physical entities, which have intrinsically only quantifiable properties and whose activities are determined only by the laws of nature, at least statistically so.

One way to think of the agreement between many (even if not all) dualists and materialists in regard to mechanism is in terms of the shared belief that *materialism gets at least most things right.* That is, our account of nonliving things, plants, and lower animals is no different whether one accepts materialism or a dualism about persons; that is, either way the mechanist takes it that the whole story about nonliving things, plants, and lower animals can be given in terms of matter, the physical properties of matter, and the relevant physical laws. The dualist, however, makes an exception in the case of humans, and maybe some higher animals, whereas the materialist thinks that mechanism gets *everything* right.[7] Given this understanding of materiality, it is not hard to see why dualism and materialism, though they are contrary positions, are in many ways complementary. If matter essentially excludes the qualitative and intentional aspects that are the mark of psychological properties,

5. *The Catholic Encyclopedia,* accessed on 8/20/10 at http://www.newadvent.org/cathen/10100a.htm.

6. Unger, *All the Power in the World,* 19.

7. Rene Descartes, the most influential modern dualist, gives a mechanistic account of what *seems* to us to be sensation in nonhuman animals in *Discourse on Method,* part 5, in *The Philosophical Writings of Descartes* vol. 1, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch, 138–41 (New York: Cambridge University Press, 1985).

then it seems that one who believes that there are psychological properties will be required to conclude that they occupy a nonspatial, nonextended ontological realm. That would be to embrace some sort of dualism. On the other hand, one might either deny the reality of psychological properties altogether or attempt to show how psychological properties can turn out to be something quantifiable, and so on, despite the seemingly obvious appearance otherwise. That would be to embrace some sort of materialism. In short, it seems that if matter is best understood mechanistically, then one will tend to embrace either materialism or dualism. As we discuss the arguments for dualism in this chapter, this point will become all the more clear.

For many people, dualism enjoys a great deal of intuitive plausibility; it seems to be obviously true that sensations and thoughts are fundamentally different events or processes from the growth of a blade of grass or the falling of a heavy object to the earth. I do not mean to beg the question at the onset by claiming that common sense lends itself to some form of dualism. My point is merely that prior to sophisticated philosophical reflection it seems to many people as though sensations and thoughts are not physical phenomena, but that is not to assert that dualism is true. There was a time when the idea that the sun revolves around the earth likewise enjoyed a great deal of intuitive appeal and stood as a pillar of common sense, and we know now that this idea is very wrong. In this light, the prima facie tendency toward dualism is granted by some of the most strongly committed critics of dualism and adherents of materialism.[8] For example, Richard Rorty, himself a materialist of the strictest variety, admits that

> we seem to have no doubt that pains, moods, images, and sentences which "flash before the mind," dreams, hallucinations, beliefs, attitudes, desire, and intentions all count as "mental" whereas the contractions of the stomach which cause the pain, the neural processes which accompany it, and

8. Charles Taliaferro gives references to several physicalists who make this concession in *Consciousness and the Mind of God* (Cambridge: Cambridge University Press, 1994), 26.

everything else which can be given a firm location within the body count as nonmental. Our unhesitating classification suggests that not only have we a clear intuition of what "mentality" is, but that it has something to do with nonspatiality and with the notion that even if the body were destroyed the mental entities or states might somehow linger on.[9]

Given the fact that dualism enjoys a common, intuitive appeal, one might be inclined to conclude that the burden of proof rests squarely on the shoulders of the materialist. That is, one might argue that our default position should be dualism, and it is therefore up to the materialist to prove that dualism is wrong just as the prosecution in a court of law carries the burden of proof. If the prosecution fails to prove its case, then the accused goes free; likewise, if the materialist fails to prove dualism is false, then dualism is vindicated.

Appeals to the burden of proof frequently cut both ways, and in the end they are often quite unconvincing. For example, Kevin Corcoran, himself no naturalist, finds materialism to be intuitively plausible, and dualism almost unthinkable:

Try as I might, I cannot bring myself to believe what my mother believes, and what many Christians down through the ages have believed, about the metaphysics of human persons. It is not that I do not understand the view. I do. So too with the traditional way of understanding the nature of human persons. I simply cannot believe that I am an immaterial thing. I can believe that some kinds of persons are immaterial—for example, nonhuman, divine persons like God and the angels. But human persons like me. That I cannot believe.[10]

J. J. C. Smart cites a similar intuition at the outset of an influential defense of materialism:

So sensations, state of consciousness, do seem to be the one thing left outside of the physicalist picture, and for various reasons I just cannot believe that this can be so. That everything should be explicable in terms of physics

9. Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton, N.J.: Princeton University Press, 1979), 17.

10. Kevin Corcoran, "The Constitution View of Persons," in *In Search of the Soul: Four Views of the Mind-Body,* ed. Joel B. Green and Stuart L. Palmer (Downers Grove, Ill.: InterVarsity Press, 2005), 154.

… except the occurrence of sensations seems to me to be frankly unbelievable.[11]

There is a place in philosophy for worries about the burden of proof, and we will indeed return to this issue later in our discussion. For the moment, however, we will leave such considerations aside, since there are appeals to common sense on both sides. We will assume hereafter that materialism and dualism are both prima facie plausible theories; that is, we have no overridingly good reasons to believe that either one of them is preferable without first investigating the arguments their proponents give in their favor and the arguments given by their opponents in order to refute them. In this chapter we will focus on the arguments for dualism, and in the next chapter we will consider the case against dualism.

Before turning to the most prominent arguments for dualism, we will pause to consider one common set of arguments attempting to show that the project of defending substance dualism at all is completely misguided, though they would leave property dualism unscathed. According to the proponents of this position, there is no need to investigate the arguments for substance dualism, because the very notion of a nonphysical substance is either vague or even incoherent. Since there is no reason to listen to arguments purporting to prove an incoherent conclusion, we should then save ourselves the trouble of looking into arguments for dualism. In addition to claiming that a mind would be the nonphysical bearer of nonphysical, psychological properties (per the substance dualist), I have had little to say as to *what such a nonphysical substance is* (as opposed to what it is not), and it is notoriously difficult to do otherwise. One might wonder along with Elizabeth Anscombe (herself no materialist) whether "the conception of an *immaterial substance* at all—is a delusive one."[12] An-

11. J. J. C. Smart, "Sensations and Brain Processes," in *Materialism and the Mind-Body Problem,* ed. David M. Rosenthal (Indianapolis, Ind.: Hackett Publishing, 2000), 55.

12. Elizabeth Anscombe, "The Immortality of the Soul," in *Faith in a Hard Ground: Essays on Religion, Philosophy, and Ethics by Elizabeth Anscombe,* ed. Mary Geach and Luke Gormally (Exeter, UK: Imprint Academic, 2008), 71.

scombe's point is that our understanding of "substance" is inextricably linked to physical objects. On her view, it is only through encountering tangible physical things that we come to grasp this concept, and its use is thereafter always parasitic on the reference to such objects. To claim, as does the substance dualist, that we can understand the idea of a substantial, nonphysical mind is really to fail to make sense, because any such claim betrays only a misunderstanding of the meaning of "substance."[13]

Furthermore, difficulties in articulating what would differentiate one nonmaterial substance from another lead some philosophers, most prominently in recent decades Anthony Flew, to reject the very notion of a nonphysical substance as unintelligible.[14] Colors, shapes, locations, and the like differentiate one physical object from another. For example, two highly similar dogs, Fido and Rex, are nevertheless differentiated by even slight differences in their size, shape, color, and so on; for example, Fido has a slightly darker coat than Rex. Even if Fido and Rex were exactly the same size, color, shape, and so on, they could still be differentiated because they occupy different positions in space. None of these properties, however, can differentiate nonphysical substances, because minds are supposedly nonspatial, colorless, shapeless, and so on. It is quite mysterious then as to what differentiates two numerically distinct, nonphysical substances. If the very notion of a nonphysical substance were intractably vague, or even incoherent, there would be little reason even to consider the arguments in favor of substance dualism.

I do not deny that it is very difficult to articulate what we mean by "nonphysical substance," and life for the substance dual-

13. Ibid., 71–72.

14. See Anthony Flew, "Death," in *New Essays in Philosophical Theology,* ed. Antony Flew and Alasdair MacIntyre, 267–72 (London: SCM Press, 1955), and *God and Philosophy* (Amherst, N.Y.: Prometheus Books, 2005), 46. Flew now seems to have abandoned this objection with respect to the possibility of disembodied minds in general, e.g., God or angels, but it is unclear whether he believes that human minds can be disembodied. See Flew's *There Is a God: How the World's Most Notorious Atheist Changed His Mind* (New York: Harper Collins, 2007), 147–54.

ist would certainly be easier if we could do so. I agree with Anscombe when she claims that our understanding of "substance" is something we first gain in our dealings with tangible objects (all of which are physical objects), but this is a controversial philosophical thesis that would need to be defended at great length. Rene Descartes, among others, argues vigorously that we derive our initial notion of *substance* from our direct awareness of ourselves as nonphysical substances, and only subsequently can we apply the notion of *substance* to physical objects.[15] I do not, however, wish to enter into that debate, so let's just grant Anscombe this premise for the sake of argument.

This concession, nevertheless, need not be a deal-breaker for substance dualism. First of all, the fact that we lack an informative definition of "nonphysical substance" is not a strike against the substance dualist, because our concept of "physical substance" is no better off (at least for the mechanist). What it is for something to be material or physical consistent with mechanism, whether it is a property or a substance, is a matter of contentious philosophical debate, for which no resolution is in sight. Indeed, this seems to be not a matter of mere disagreement, but an apparent case of our just not having a clear-cut sense of what we mean by "physical" at all in the mechanist's sense: as Bertrand Russell famously puts it, "we know nothing about the intrinsic quality of physical events except when these are mental events that we directly experience."[16] Moreover, the identity and differentiation of physical objects is one of the perennial problems of philosophy, regard-

15. See Rene Descartes, *Meditations on First Philosophy,* in *The Philosophical Writing of Descartes,* vol. 2, trans. J. Cottingham, R. Stoothoff, and D. Murdoch (Cambridge: Cambridge University Press, 1985), 16–23

16. Bertrand Russell, "Mind in Matter," in *Portraits from Memory* (Nottingham: Spokesman, 1956), 153. Charles Taliaferro enumerates seven different ways of defining "physical," all of which are subject to reasonable criticism, in *Consciousness and the Mind of God,* 90–103. Sandra Menssen and Thomas Sullivan also highlight the difficulties involved in defining "physical" in *The Agnostic Inquirer: Revelation from a Philosophical Standpoint* (Grand Rapids, Mich.: Eerdmans Publishing, 2007), 40–45. Galen Strawson also argues that we do not possess a rigorous conception of the intrinsic nature of the physical in *Real Materialism and Other Essays* (New York: Oxford University Press, 2008): 20–26.

ing which there is no consensus view even among contemporary philosophers.[17] Thus, when asked, "What is a physical object?" unless we are willing to beg important philosophical questions, it seems that our best answer is "I don't know exactly, but I know one when I see one." There is no problem with such an answer as long as we are willing to admit that our position presupposes a "promissory note" in the metaphysics of material objects or that the notion of a physical object is a first principle that cannot be explained in any more fundamental terms. Dean Zimmerman makes this point when addressing a similar objection regarding the criteria for identity of nonphysical substances:

> How do we know that God, or quantum-mechanical whimsy, is not playing similar tricks on the physical plane? Perhaps my body is periodically annihilated, replaced by a duplicate so quickly as to fool even the most careful observer. The right response is, surely, that although it's possible, it's not something anyone should lose sleep over. But then why can't the dualist say the same thing?[18]

In other words, if we are willing to allow promissory notes or appeals to disputed first principles to bolster a notion of physical substance, I see no problem in likewise allowing similar maneuvers for the dualist with respect to the notion of nonphysical substance. The substance dualist may admit that she cannot give a clear, positive definition of "nonphysical substance," including criteria for identity and differentiation, but she nevertheless knows one when she "sees" one.

Secondly, our notion of substance may well be primarily formed through our encounter with the physical objects of ordinary experience, but we are often able to extend our concepts beyond their original experiential application in order to make sense of things of which we have direct experience. John Haldane offers the example of inferring certain facts about the blockage

17. See Van Inwagen, *Metaphysics,* 23–27, and Lowe, *A Survey of Metaphysics,* 23–78.

18. Dean Zimmerman, "Christians Should Affirm Mind-Body Dualism" in *Contemporary Debates in Philosophy of Religion,* ed. Michael L. Peterson and Raymond J. Vanarragon (Malden, Mass.: Routledge, 2004), 319.

in a hot water heater from the lack of warm water in his morning shower.[19] He has no direct experience of the blockage, but he postulates such an entity with certain properties based on what is necessary (or most likely) to explain certain facts he does experience, namely the lack of warm water. It may even be that doing so requires him to make novel applications of previously held concepts. Maybe Haldane has never experienced the sort of blockage necessary to hamper a hot water heater, but he could nevertheless apply his prior concepts of physical objects and relations to speculate reasonably about the cause of the lack of warm water for his morning shower. It is in this way that we posit the existence of unseen entities, such as electrons and quarks, in scientific theories; they are the best available explanations of the phenomena we are able to observe, even though they are beyond the boundaries of direct experience. Moreover, positing theoretical entities such as electrons often requires us to extend our application of previously understood concepts, for example, "particle" as applied to electrons, in new ways. Our concept of particle originates in our empirical encounters with objects such as specks of dust, which have very little in common with electrons. We have no problem, however, extending the notion of particle beyond its original strict application to include electrons. Likewise, the dualist might argue, we have no experience of nonphysical substances (although some dualists following Descartes would doubt this), and "substance" is certainly more at home when applied to tangible physical objects, but things of which we do have experience, for example, sensations and thoughts, lead us to infer that there are nonphysical substances to provide their most plausible explanation. In making this inference we are extending the concept of substance beyond its normal use, but there is nothing particularly incoherent in doing so. The dualist will argue that we know nonphysical substances through their effects, and we may

19. John Haldane and J. J. C. Smart, *Atheism and Theism* (Malden, Mass.: Wiley-Blackwell, 1996).

therefore extend the use of certain concepts to characterize such entities. We will assume hereafter that substance dualism is not intrinsically incoherent or unintelligible, and now turn to the most common forms of argument in its favor.

The Modal Argument for Substance Dualism

In the remaining sections of this chapter we will discuss three of the most influential of arguments for dualism.[20] All of these arguments have a rather long history, and they have been variously developed over the centuries. My intent is not to reproduce these arguments as they appear exactly in the writings of any particular philosopher, but to give a presentation of the general sweep of these lines of reasoning, which will at times require a fair degree of simplification. The argument we discuss in this section originates in the work of the seventeenth-century philosopher, scientist, and mathematician, Rene Descartes, whose work is generally taken to mark an epochal change in the history of philosophy, a transition from the medieval to the early modern period of thought. This certainly greatly simplifies the complex turning point in the history of ideas that occurred in the early modern period, but it is true that Descartes is one of the most influential figures in the history of modern philosophy. Even though he is commonly maligned today (and many philosophers spend a great deal of energy doing so!), it cannot be denied that Descartes has set the stage for much of the philosophizing that has come after him, especially because Descartes is far and away modern philosophy's most influential substance dualist. In fact many philosophers refer to substance dualism as "Cartesian Dualism." Even though Descartes's views

20. This is not to say, of course, that we will discuss all of the important arguments for dualism in this chapter. In fact, I have consciously made some important omissions. Namely, what are sometimes called "qualia arguments," "knowledge arguments," or even "zombie arguments" are among the most influential arguments for dualism on the contemporary scene, but I have left them for chapter 5. I do so because these arguments are likely more easily understood after the reader has had some exposure to the sort of materialism their proponents reject. Please note, however, that these arguments not only are critiques of materialism, but likewise constitute significant positive cases for dualism.

do not enjoy widespread acceptance among mainstream professional philosophers in Western universities today, many of those who do defend substance dualism do so based on arguments that originate with Descartes.[21]

Before we look at the first of the arguments for substance dualism, we need to grasp some very important concepts and distinctions. I beg the patience of readers inexperienced with the distinctions I am about to make over the course of the next few pages, as they may at first be difficult to identify as philosophically important. Please trust, however, that this bit of logical hair-splitting is going to pay off for us in nearly all of our discussions to follow. The first of these distinctions is taken from *modal logic,* which is concerned with different modes or ways a proposition can be true or a state of affairs can obtain. Consider the following proposition:

(1) Jim Madden is the father of Jack Madden.

Proposition (1) is true, as my fifth child is Jack. So far so good, but now consider a second proposition:

(2) Nobody is taller than himself.

Certainly, (2) is true, but it is true in a different *mode* or *way* from the manner in which (1) is true. Proposition (1) is true, but the world could have been such that it were false. It is *unnecessary* that Jack is my fifth child, because I might not have had a fifth child at all or my fifth child might have been a girl named Brigid. We then say that (1) is *contingently true,* that is, it is true, but it could have been false. Proposition (2), however, could not have been false. I am 5′8″ tall, and I could fail to be that height either by having an odd (maybe miraculous) midlife growth spurt or by meeting with an unfortunate accident. Though I could fail to be 5′8″, whatever height I am, I will be exactly that height and not

21. For recent defenses of substance dualism using arguments drawing inspiration from Descartes, see John Foster, *The Immaterial Self* (New York: Routledge, 1991), and Stuart Goetz, "Substance Dualism," in *In Search of the Soul: Four Views of the Mind-Body Problem,* 33–60.

another, and this is true not just of me, but of everything that has a height. Something's being a certain height and not that height simultaneously would be a logical contradiction, which can never under any circumstances obtain. Thus, nothing is taller than itself. We then say that (2) is *logically necessary,* that is, it is true and it cannot fail to be true.

What about (3)?

(3) Jim Madden is the father of two daughters.

Proposition (3) is false, because I have only one daughter. Consider (4):

(4) Jim Madden is the father of infinitely many daughters.

Both (3) and (4) are false, but they are false in different ways. Proposition (3) is false, but it could be true. Proposition (4), on the other hand, cannot be true. There is no way that things could go such that a finite being, such as myself, has infinitely many daughters. We can say that, though it is false, (3) is possibly true, whereas (4) is *impossible* or *necessarily false.*

It will be helpful at this point to introduce a common convention for talking about modality employed by contemporary philosophers. A *possible world* is a coherent way the world could be in its totality, and the *actual world* is the way the world really is. There is a possible world in which I do not write this book, because that is a coherent way things could have gone. Likewise there are possible worlds in which George Washington was not the first president of the United States of America and in which Lake Michigan is fourteen feet deeper than it is now. Basically, any total or comprehensive combination of compatible states of affairs amounts to a possible world; that is, there is a way the world could be such that all of those states of affairs obtain together. Indeed, it is sometimes argued that there are possible worlds in which the laws of nature are different from those in the actual world. For example, there are possible worlds where pigs (or at least beings very similar to pigs) can fly under their own

power or the solar system is not heliocentric. Some philosophers think of possible worlds as real, alternative universes that actually exist, whereas other philosophers believe that possible worlds are merely ways of conceiving how things might be, even if not actually so. Though these debates are immensely important and rewarding, we will not wade into such deep metaphysical waters here. Suffice to say, as long as one admits that things could be otherwise than they are, then one might use talk of possible worlds to depict these alternative states of affairs. For our purposes I will assume that a possible world just is a way the world could be, whatever that may ultimately amount to metaphysically.[22]

Return now to (1). We know that (1) is true in the actual world, but false in at least one other possible world—any possible world in which I don't have a fifth child named Jack. All contingent propositions are like that; that is, they are true in at least one possible world and false in at least one possible world. Proposition (2), however, is true in all possible worlds, because there is no way a world could be such that something exists that is taller than itself in that world. We say then that necessarily true propositions are true in all possible worlds. Truths of mathematics, for example, $2 + 2 = 4$, and logic, for example, *If A is identical to B and B is identical to C, then A is identical to C,* are certainly true in all possible worlds. In addition to truths of logic and mathematics, metaphysical necessities obtain in all possible worlds. For example, *If Jim Madden exists, then he is a human being* is true in all possible worlds, since anything that fails to be a human being would not be me, even though this is not a truth of logic or mathematics. Our (3) above is contingent, and by that we mean

22. For a good introduction to the logical and metaphysical problems surrounding talk of possible worlds, see Kenneth Konyndyke, *Introductory Modal Logic* (Notre Dame, Ind.: University of Notre Dame Press, 1986), 101–19; and Michael J. Loux's introduction to *The Possible and the Actual: Readings in the Metaphysics of Modality,* ed. Michael J. Loux (Ithaca, N.Y.: Cornell University Press, 1979), 15–65. For criticism of the use of possible worlds to think about modality, see James F. Ross, *Thought and World: The Hidden Necessities* (South Bend, Ind.: University of Notre Dame Press, 2008).

that it is true in at least one possible world and false in at least one possible world. There is a possible world in which I have two daughters, even though that world is not the actual world. Proposition (4) is false in all possible worlds; that is, there is no possible state of affairs in which I have infinitely many daughters. We say then that necessarily false propositions are false in all possible worlds, and possible propositions are true in at least one possible world.[23] These distinctions may seem both tedious and painfully obvious at the moment, but they will be immensely useful in what follows.

Speaking of painfully obvious propositions, have a look at (5):

(5) Jim Madden is Jim Madden.

This proposition is what we call *an identity statement,* as it states that something is the same things as itself. Identity in this sense doesn't mean very similar, as when we speak of identical twins. Twins are very similar, though not identical, because despite their striking similarity they are distinct entities. What we mean by identity here is *the relation everything has to itself such that it is that thing and not something else.* To say that two apparently distinct things are actually identical is to say that they are really not distinct, that is, they are the same entity. Have a look at another, less obvious, identity statement:

(6) Jim Madden is Jennifer Madden's husband.

This is an identity statement, because it states that two apparently distinct things are in fact identical, that is, the very same entity. I am in fact Jennifer Madden's husband.

The modality of identity statements like (6) can be a rather complicated affair. In general, identity statements are necessarily true; that is, they are true in all possible worlds, as is no doubt the

23. Notice that possible propositions are distinct from contingent propositions, because the former includes and the latter excludes necessary propositions. Since necessary propositions are true in all possible worlds they are true in *at least one* possible world, which means they are possible. Necessary propositions, however, are not false in any possible world, so they are not contingent.

case in (5). One might interpret (6) as though it were not really an identity statement, that is, it can be taken to mean that Jim Madden has the property of being Jennifer Madden's husband, which of course is contingent, because Jennifer might have married somebody else. We can interpret (6), however, as an identity statement, which renders it necessarily true. If we are using "Jim Madden" to refer to the man typing this paragraph in every possible world in which he exists and we are using "Jennifer Madden's husband" to refer to the man typing this paragraph in every possible world in which he exists, without concern for what these phrases say about their objects, then (6) is necessarily true. We will, for now, take it that true identity statements are all necessarily true; that is, they are true in all possible worlds, although there are some cases, such as (6), wherein this assumption depends on how we interpret the terms composing the statement.[24]

With these logical distinctions and a basic understanding of identity in hand, we may now turn to what is called the "modal argument for substance dualism" (hereafter just the "modal argument"). The following argument is inspired by (though not a strict reconstruction of) a very sophisticated line of reasoning that Descartes presents in the "The Sixth Meditation" of his *Meditations on First Philosophy,* an argument that has subsequently been developed by various philosophers over the years.[25]

24. We will discuss the notion of so-called *contingent identity* (though *empirical identity* would be a better way to put it), which I have set aside here for the sake of letting the dualist arguments get off the ground, in the following chapter.

25. For Descartes's original presentation of the argument see Rene Descartes, *Meditations on First Philosophy,* 54. For interpretations of Descartes's version of the modal argument see Georges Dicker, *Descartes: An Analytical and Historical Introduction* (New York: Oxford University Press, 1993), 188–99. Joseph Almog offers an interpretation and critical analysis in his *What Am I? Descartes and the Mind-Body Problem* (New York: Oxford University Press). The contemporary literature on the modal argument is extensive, and I make no pretense to addressing it here. The most influential contemporary version of the modal argument is Saul Kripke's *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1980), 144–55. Christopher Hill provides a good overview of Kripke's position and the issues in the associated literature, along with an incisive critique in "Imaginability, Conceivability, Possibility, and the Mind-Body Problem," *Philosophical Studies* 87, no. 1(1997): 61–85. Keith Yandell offers a defense of the modal argument in "A Defense of Dualism," *Faith and Philosophy* 12, no. 4 (October 1995): 548–66. See also Taliaferro,

(1) If materialism is true, then my psychological states are identical to processes or events in my brain, call them collectively *B*. (premise)

(2) If my psychological states and *B* are identical, then necessarily my psychological states and *B* are identical. (premise)

(3) It is possible that my psychological states exist without *B*. (premise)

Therefore:

(4) It is not necessary that my psychological states and *B* are identical. (from 3)

Therefore:

(5) My psychological states and *B* are not identical. (from 2 and 4)

Therefore:

(6) Materialism is false. (from 1 and 5)

(7) Either materialism or dualism is true. (premise)

Therefore:

(8) Dualism is true. (from 6 and 7)

Premises (1), (2), and (3) are assumptions not derived from other premises in the argument. (1) is plausible, since the materialist claims that psychological states are physical realities just like any other part of or event in the human body. Thus, if materialism is true, then presumably our psychological states are identical to some part or process of the brain, what I'm calling *B,* because we know, both scientifically and as a matter of common sense, that psychological states are intimately connected to states of the central nervous system. Thus, the prospects for materialism are presumably tied to the prospects of arguing that our psychological states are identical to *B.*[26]

Premise (2) may initially seem curious, because it implies that

Consciousness and the Mind of God, 173–88; and W. D. Hart, *The Engines of the Soul* (Cambridge: Cambridge University Press, 1988).

26. There are versions of materialism that actually deny (1). For example, *eliminative materialists* deny that there are any such things as psychological states, so psychological states are not identical to *B,* because there are no such states. For the moment, we will presume a simple version of materialism that does identify psychological states with *B,* though alternative versions of materialism will be considered in great detail in chapter 4.

necessity can be inferred from actuality, which in most cases is absurd. For example, "Will is playing in the yard" certainly does not imply that "Necessarily, Will is playing in the yard." Just because Will is playing in the yard in the actual world does not imply that Will is playing in the yard in all possible worlds. Thus, in ordinary cases involving contingent propositions, actuality does not imply necessity. Identity statements, however, behave a bit differently. As we discussed above, if an ordinary identity statement is true, it is necessarily true. If William is identical to Patrick in the actual world, then William is identical to Patrick in all possible worlds, given the qualifications above; that is, we assume that the references of "William" and "Patrick" remain constant across possible worlds. Thus, if my mind and my brain are identical, then they are necessarily identical. Though on the face of it (2) may look strange, upon reflection it should not worry us.

Premise (3) is going to be a point of controversy, so we will set it aside for the moment. We can get premise (4) easily from premise (3). To say, as in (3), that one entity could exist without the other, is to imply that there is a possible world in which the one exists without the other. The proponent of the modal argument claims that there is a possible world in which my psychological states exist without *B* or presumably any other brain state. In such a possible world, the psychological states and *B* would not be identical, because only the psychological states exist in that possible world. Thus, my psychological states and *B* are not identical in all possible worlds, which is to say, as is claimed in (4), they are not necessarily identical.[27]

Premise (5) follows from (2) and (4). Suppose we know that "If it is sunny, then Patrick will go fishing" and "Patrick will not go fishing." We may then infer that "It is not sunny," because if it

27. In fact we could further infer that the mind and the body are not identical in any possible world, given premise (2). That is, if we accept (2), then we may conclude that the mind and the body are necessarily not identical. We can leave this point aside, as our version of the modal argument doesn't require these stronger implications.

were sunny, Patrick would go fishing, which we know is not the case. Likewise, if we know that "If my psychological states and *B* are identical, then necessarily my psychological states and *B* are identical" and "It is possible my psychological states exist without *B,*" then we may infer that "My psychological states and *B* are not identical." To assert that "Necessarily my psychological states and *B* are identical" implies that in any possible world in which my psychological states exist, *B* must exist in that world also, because they are the same collection of entities. If my psychological states could exist without *B,* then there must be a possible world where this is the case. Thus, if we assume that my psychological states can exist without *B,* then it follows that my psychological states are not identical to *B,* which is our premise (5).

Premise (6) obviously follows from (1) and (5). Since the truth of materialism as we are construing it at this point in our discussion implies that my psychological states and *B* are identical and we now find in (5) that my psychological states and *B* are not identical, we may conclude that materialism is false. We *supposed* earlier that our only choices are between some brand of dualism or materialism, so we can at least provisionally accept (7).[28] Given that we now have reason to conclude that materialism is false, it follows that some version of dualism is true. This conclusion leaves open the question of whether the modal argument delivers either substance or property dualism. Some defenders of similar arguments have argued that the occurrence of instances of nonphysical properties requires a nonphysical substance that bears them,[29] but the possibility that there might

28. The dichotomy between dualism and materialism, as stated in (7), is likely a false one. Indeed, one of the major concerns of this book is to defend a position that is neither materialism nor dualism. We will discuss several such positions in chapters 6–8. For now, however, we will operate under this assumption simply to get dualism off the ground, though we will later put it under serious scrutiny.

29. Descartes himself believes that he has independently demonstrated that he is a thinking substance. See Descartes, *Meditations,* 18. John Foster argues that something along the lines of the modal argument requires a nonphysical substance in "A Brief Defense of Dualism."

be instances of psychological properties without any substance (material or immaterial) that has them is explicitly entertained by such philosophers as David Hume, which would seem to undermine any inference of a nonphysical substance from the occurrence of nonphysical properties.[30] A serious consideration of these thorny metaphysical issues would take us off our path, so I will leave it as an open question as to whether the modal argument, if successful, delivers property or substance dualism.

One may wonder whether property dualism is at all plausible, at least for somebody who grants a mechanistic understanding of nature. According to a mechanist, the physical particles that compose the body of a conscious organism (like any other physical substance) are supposed to have only quantifiable properties such as mass, charge, positions, velocity, to the exclusion of the qualitative and intentional aspects of psychological properties. It is, therefore, highly counterintuitive, and maybe plainly incoherent, that a mechanistically construed physical substance could exist in psychological ways (including qualitative states such as sensations).[31] That is, for the mechanist, it is difficult to understand how a physical substance could have psychological properties (whether or not psychological properties are nonphysical), because mechanism excludes everything but the quantifiable from physical things. Substance dualism does not face such a problem, because the substance dualist attributes nonphysical properties only to the mind, which is supposed to be a nonphysical substance. A property dualist might reply that the arguments for dualism simply show that some physical things have nonphysical properties, and we must therefore either revise the mechanistic

30. See David Hume, *Treatise on Human Nature* (New York: Oxford University Press, 1978), bk. 1, pt. 4, secs. 5–6. A recent defense of such a view can be found in Peter Simons, "Farewell to Substance: A Differentiated Leave-Taking," in *Form and Matter,* ed. David S. Oderberg, 22–39 (Malden, Mass.: Blackwell, 1999).

31. Here I distinguish between abstract, nonphysical properties, e.g., *being prime,* and what the dualist supposes to be nonphysical psychological properties, e.g., *believing that the sky is blue.* Following Keith Yandell in "A Defense of Dualism," 549, I see no difficulty in physical substances having the former, whereas the latter pose problems.

presuppositions of modern philosophy of mind or settle for the fact that we cannot explain how these properties can be related to our understanding of physical substances in general.[32] This is certainly true, but the fact that we cannot really see how our theory of mind (in this case property dualism) connects with our theory of the physical world (mechanism) certainly counts against the plausibility of such views. It is likewise detrimental to a theory of mind if it calls on us to revise fundamentally a general account of physical reality we seemingly hold for good independent reasons. Of course these considerations do not demonstratively refute property dualism, but only show that substance dualism has some advantages.[33] For our purposes, we can leave this question open, and simply take it that the modal argument, if successful, shows that some version of dualism is true.

Of course none of the foregoing considerations matter, unless the proponent of the modal argument can defend premise (3), that is, "It is possible that my psychological states exist without my brain." Proponents of the modal argument typically argue for (3) based on the supposed fact that we can *conceive* of psychological states without bodies. It is often proposed that possible worlds are best defined in terms of what is conceivable. Conceiving here does not mean imagining or picturing a certain state of affairs, as I can imagine a leprechaun or the older brother I have

32. Thanks to Patrick Toner for raising this objection on behalf of the property dualist. In chapter 6 we will discuss David Chalmers, Galen Strawson, Thomas Nagel, Colin McGinn, and John Searle (though they do not all identify themselves as property dualists) among other contemporary philosophers who recommend that we revise our understanding of physical reality so as to accommodate apparently irreducibly mental phenomena or admit that we cannot align certain psychological facts with mechanism.

33. I have, admittedly, been rather swiftly dismissive of property dualism. The reader interested in seeing property dualism get a full hearing in a fair court should consider Michael Tooley's fascinating defense of a very sophisticated version of property dualism in his contribution to Alvin Plantinga and Michael Tooley, *Knowledge of God* (Malden, Mass.: Blackwell, 2008), 189–205. See also Timothy O'Connor, "Causality, Mind, and Free Will," in *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons,* ed. Kevin Corcoran, 44–58 (Ithaca, N.Y.: Cornell University Press, 2001). Peter Strawson offers a highly influential version of a related view, dual aspect theory, in *Individuals: An Essay in Descriptive Metaphysics* (New York: Routledge, 1964), 87–116.

never had. Rather, to conceive of a state of affairs is to understand it; that is, we can recognize it as a logical possibility. I cannot conceive of a five-sided square (that would be impossible), but I can conceive of a *chiliagon,* a thousand-sided geometric figure. Such a figure is a logical possibility, and I can recognize it even though I cannot picture it.[34] Descartes believes that he can clearly conceive of his mind existing without his body, because it is possible for him to doubt that his body exists (it may be that our bodily lives are part of an elaborate dream or a hallucination brought on by a malicious demon), whereas he cannot coherently doubt the existence of his psychological states entirely, because doubting is itself a psychological state.[35] Thus, Descartes believes that he can conceive of his mind (the substance that has his psychological states) without his body, and therefore it is possible for his mind to exist without his body. In other words, because it is conceivable that my mind exist without my brain, there is some possible world in which my mind exists and my body does not.

The problem for (3) is that claims of conceivability are notoriously difficult to evaluate. How could somebody know whether her supposed conception of some state of affairs is legitimate or merely the appearance of being conceivable? How long does one need to think about a supposed state of affairs before he can be sure that it is in fact conceivable or inconceivable? Of course, if a state of affairs is logically contradictory, then it is inconceivable. But does that mean that we must say that all states of affairs free of logical contradictions are conceivable? Cases like "Jim Madden is a reptile" are not logically contradictory, but it's hard to say that they are conceivable. Nobody seems to have a very good answer to these questions and many others that arise regarding conceivability. Moreover, the general connection between conceivability (even if we knew when something were in fact conceivable) and logical possibility is equally contentious. It is not that there is any

34. Descartes uses this example at *Meditations,* 50–51.
35. Ibid., 16.

overwhelmingly good reason to doubt that conceivability is sufficient for logical possibility, but there isn't any overwhelmingly good reason to accept this claim either. Thus, whether somebody accepts (3) will depend not on her views specifically in the philosophy of mind, but on broader philosophical views regarding conceivability and its connection to possibility. These issues are contentious, if not outright intractable, so there is little hope in giving a knock-down argument either for or against (3) at this point.[36]

Even if we accept conceivability as a criterion for logical possibility, we have seen earlier in this chapter that some materialist philosophers claim that it is inconceivable that they exist without their bodies. We are left comparing contradictory claims of conceivability and inconceivability. This controversy may well be interminable, because we have no universally acceptable way of deciding the conceivability of states of affairs that are not outright logically contradictory. Of course various fantasy tales and Hollywood movies invoke scenarios involving persons switching bodies or existing as ghosts. There could, however, be tales about three-sided squares, but that certainly doesn't imply that such a state of affairs, which is clearly impossible, is really *conceivable* in the narrow sense relevant to the modal argument. As William Hasker puts it in his critique of the modal argument, "we can construct thought experiments in which persons are not separable from their bodies as well as thought experiments in which they are sepa-

36. See Almog, *What Am I?* 42–58 for a discussion of issues regarding conceivability, possibility, and the modal argument. See also Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), 46–49; David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press, 1996), 136–38; and Frank Jackson, *From Metaphysics to Ethics* (New York: Oxford University Press, 1998), 67–74. See also Peter van Inwagen, "Ontological Arguments," *Noûs* 11, no. 4 (1977): 375–95 for a critical discussion of appeals to conceivability in defense of possibility claims in the context of theistic arguments. For further defenses of imaginability or conceivability as criteria for possibility, see Taliaferro and Evens, *The Image in the Mind*, 224–28, and Taliaferro, "Possibilities in the Philosophy of Mind," *Philosophy and Phenomenological Research* 57, no. 1 (1997): 127–37. For a defense of our ability to grasp modal facts, see Timothy O'Connor, *Theism and Ultimate Explanation: The Necessary Shape of Contingency* (Malden, Mass.: Blackwell, 2008).

rable."[37] Thus, the proponent of the modal argument claims that she can conceive of her mind without her body, while the materialist claims she cannot conceive nonbodily existence, and we are seemingly left with no principled reason to accept either claim.

Peter Geach has argued against (3) not only by appealing to a counterintuition regarding conceivability, but also by arguing that we cannot separate the concept of sensation from other concepts that imply physical embodiment. Take the example of the concept of *seeing.* Geach argues that our understanding of *seeing* is derived from certain nonpsychological concepts, for example, tangible physical objects and sense organs. I cannot separate, says Geach, "seeing" from "seeing something at a certain distance away" or some other spatial association. The notion of an entirely nonspatial seeing, for Geach, is unintelligible; we literally don't know what we mean by "seeing" in such a circumstance: "the concept of seeing can be maintained only because it has threads of connexion with these other nonpsychological concepts; break enough threads, and the concept of seeing collapses."[38] "Seeing," like all sensation terms, is inextricably bound up with our physical existence. Thus, the very notion of a disembodied experience is impossible, says Geach, and therefore (3) is false. In other words, Geach replies to Descartes by arguing that in any world in which his mind exists without his body, none of our usual sensation concepts would be available, because these concepts are dependent on a context of physical objects for their proper application. It is then difficult to see how a proponent of the modal argument can reasonably claim to have a conception of himself as a disembodied mental substance, since anything we can intelligibly recognize as having sensations would be unavailable in any such world.[39]

37. William Hasker, *The Emergent Self* (Ithaca, N.Y.: Cornell University Press, 1998), 118.

38. Peter Geach, *God and the Soul* (South Bend, Ind.: St. Augustine Press,1969), 21. Wilfrid Sellars famously makes a similar point in "Empiricism and the Philosophy of Mind," in *Science, Perception, and Reality* (Atascadero, Calif.: Ridgeview Publishing, 1991), 140–49.

39. I am leaving aside the issue of whether thoughts should be treated separately

So far things don't look great for (3); it seems that at best the case for (3) is inconclusive because it faces the challenge of contradictory inconceivability claims that cannot be easily answered, and at worst (3) is shown to be false because psychological concepts presuppose a material context for their significance. Things may not, however, be so bad for (3) after all. You are likely familiar with (hopefully by way of secondhand information) what are commonly called *near-death experiences,* NDEs, wherein people who have been pronounced clinically dead following a complete cessation of respiration or detectable brain function seemingly continue to have experiences that they recall after being resuscitated, or, some would speculate, after returning from the dead. Sometimes these experiences involve mystical visions, glimpses of an afterlife, and a review and judgment of one's life. In other cases, NDEs involve the ability of patients to recall conversations had by their doctors and even events that occur far away from the location of the patient's body.[40]

Some philosophers make much of NDEs as evidence for dualism; if consciousness continues even after the body (including the central nervous system) has succumbed to death, it would seem that the mind is a separate substance from the body and can exist on its own.[41] My own view is that NDEs do not constitute much of a direct case in favor of substance dualism, or even personal immortality for that matter. It seems that hallucinations brought on by a dwindling supply of oxygen to the brain (which

from sensations in arguments against materialism or for dualism. We will take this issue up in some detail in chapter 5 below. It is, however, important to keep this distinction in the back of your mind. For instance, though Geach doubts whether sensations have much to show us about the possibility of disembodied existence, he argues elsewhere that thoughts give us some reason to believe that the mind, in some sense, can survive bodily death. See Geach, *God and the Soul,* 30–41.

40. For a presentation and analysis of many cases of apparent NDEs see Raymond Moody Jr., *Life after Life* (Atlanta, Ga.: Mockingbird Books, 1975) and Michael Sabom, *Recollections of Death: A Medical Investigation* (New York: Harper & Row, 1982).

41. See chs. 8–9 of Gary Habermas and J.P. Moreland, *Beyond Death: Exploring the Evidence for Immortality* (Wheaton, Ill.: Crossway Books, 1998). Charles Taliaferro uses reports of out-of-body experiences (but not NDEs specifically) to defend a premise similar to our (3) in *Consciousness and the Mind of God,* 175–76.

one might expect under near-death conditions), lingering brain activity not detectable by our current instrumentation, misplaced memories, and a great many other plausible explanatory scenarios are at least as likely explanations as substance dualism.[42] Nevertheless, even if one denies, as I am inclined, that NDEs are literally cases of disembodied experience, the fact that they seem to be coherent as experiences might be enough to support (3). As Taliaferro and Evans put it, "But even granting that [NDEs] are false or unreliable, don't they at least appear to describe a coherent, bona fide possibility?"[43] Of course Geach would argue that NDEs are not really coherent experiences (somebody might imagine having such an experience, but she can't imagine the experience coherently), but the way NDE skeptics address these purported NDEs gives us some reason not to rule them out as incoherent. Notice that typical reasons for rejecting NDEs do not proceed from a presumption of the inconceivability of disembodied experience. When someone claims to have had a NDE, our reply is typically not to point out to him that he is obviously mistaken because such experiences are logically impossible or flatly inconceivable. Rather, we tend to reply by offering other explanatory hypotheses for how the experience may have come about, and we leave it up to empirical study to determine which is better confirmed. If somebody proposed to have seen a man who was both 6′7″ and not 6′7″, discovered something that is literally both a horse and a lizard, or some other inconceivable state of affairs, we would not bother attempting to refute him empirically. The point is that if disembodied existence were inconceivable, then we should reject claims to have had such experiences without bothering with any empirical investigation. That is not the typical reply to NDEs, even among materialists. It would

42. See Susan Blackmore, *Dying to Live: Near-Death Experiences* (Buffalo, N.Y.: Prometheus Books, 1993), for a discussion of various alternative hypotheses for explaining NDEs. Habermas and Moreland reply to some of Blackmore's arguments in *Beyond Death*, 206–10.

43. Taliaferro and Evans, *The Image in Mind*, 186.

seem then that disembodied experience, though it may not ever happen, is not strictly inconceivable, and this fact weighs in favor of premise (3) of the modal argument.

As we have seen, Geach argues to the effect that at least our concepts of sensations are inextricably linked to material reality, because these concepts are properly used only when we are referring to material objects, or at least it is our encounter with such objects that accounts for our original conceptual acquisition. We addressed an argument similar to Geach's when we discussed an objection to the intelligibility of the notion of a nonphysical substance, and we could extend replies to that objection to cover Geach's worries.[44] That is to say, one can grant Geach his account of how mental concepts are obtained, but then point out that his position does not preclude our applying such concepts to novel situations when we have very good theoretical reasons for doing so. We stretch our common, experiential concepts to their very limits when doing natural science, and there seems to be no good reason to think we cannot do so when wondering about the nature of the mind. Thus, just because our mental concepts originally apply in situations wherein we are involved with physical objects, it does not follow that we can never use such concepts to refer to minds in possible worlds wherein there are no material objects. Moreover, part of Geach's claim seems to be that in the actual world we get our mental concepts through a process that presupposes material objects, including our bodies (brains), so it is incoherent to say that we have such concepts in possible worlds in which our bodies do not exist. Geach, however, ignores the possibility that in those other possible worlds there could be other ways of obtaining mental concepts. It is conceivable that, pace Descartes, some mischievous demon gives us those concepts by some sort of direct illumination (dis-illumination), or even that these concepts are innate. Even if we are not willing to entertain the possibility of a mischievous demon or innate ideas,

44. See the first section of the current chapter.

such scenarios are still conceivable (as far as we can tell). There is, it seems, very little reason to believe that disembodied experience is inconceivable given Geach's argument alone.[45]

So where does the modal argument stand? The crucial premise is (3), and there is something to be said in its favor. However, one may not take the conceivability of NDEs as a particularly strong piece of evidence for the key premise in an immensely important philosophical argument.[46] We have no reason to believe that disembodied experience is inconceivable, and some, maybe slight, reason to believe that it is conceivable. Even if we accept the slight favor that the plausibility of NDEs gives to premise (3), there is still the general philosophical worry about whether conceivability is really a sound criterion for determining possibility. In the end, we may do best to remain sympathetic to, yet ultimately agnostic about, (3), which likewise requires us to withhold judgment regarding the argument in general. The modal argument, though it shows that dualism is not completely without merit, nevertheless leaves the issue an open question.[47]

The Difference Argument for Dualism: Unity and Simplicity

Leaving aside the modal argument, we now turn to what I will call the *difference argument.* Descartes certainly offers an influential version of the difference argument, but a great many seminal thinkers in the history of philosophy who deny material-

45. I should note that I am generally in agreement with the epistemic priority that Geach, along with Sellars, gives our understanding of physical objects over our understanding of inner episodes such as sensations. To take Sellars's way of making the point, I agree that inner episodes (sensations and thoughts) are posited as part of a theory aimed at explaining events and activities in the outer, material world. What I deny is that once we have such a theory, we cannot extend its applications to include scenarios that might even include disembodied existence. We will return to related considerations in subsequent chapters.

46. William Hasker doubts that the conceivability of NDEs is at all helpful for the modal argument. See Hasker, *The Emergent Self,* 117–22.

47. We will be able to level some more sophisticated criticism of the modal argument once we have discussed the identity theory in section 3 of chapter 4 below.

ism do so for reasons that can be broadly construed as a version of this argument. Here is a schema that encompasses all versions of the difference argument, which is a much simpler line of reasoning than the modal argument:

(1) If the mind and some part of the brain, *B,* are identical, then the mind and *B* must have all of their properties in common. (premise)

(2) The mind has a property, *F,* and *B* lacks *F.* (premise)

Therefore:

(3) The mind and *B* are not identical. (from 1 and 2)

Therefore:

(4) Dualism is true. (from 3)

Premise (1) is an application of what is commonly known as *Leibniz's Law,* named for the seventeenth-century philosopher who famously employed it. Certain versions of Leibniz's Law are quite controversial, but the version used in the difference argument should not give us too much trouble.[48] Suppose that somebody claims that Patrick and Brendan are in fact identical, but we know that Patrick has the property of having brown hair, which Brendan lacks. If Patrick and Brendan were identical, it would then follow that Patrick has and does not have the property of having brown hair at the same time, which would also be the case for Brendan. Nothing can both have and lack a property at a given moment; logically contradictory states of affairs are not possible. Thus, given the difference in properties, Patrick and Brendan are

48. We are employing what is often called the *principle of the indiscernibility of identicals,* i.e., if x is identical to y, then x and y have all of their properties in common, but Leibniz himself additionally asserts the *principle of the identity of indiscernibles,* i.e., if x and y have all of their properties in common, then x and y are identical. Some philosophers deny the latter principle because they believe that there can be objects that have all of their properties in common without being one and the same thing. There is a vast literature on this problem, but the article that first raised doubts about the principle of the identity of indiscernibles in the twentieth century is Max Black's "The Identity of Indiscernibles," in *Classics of Analytical Metaphysics,* ed. Larry L. Blackman, 281–92 (Lanham, Md.: University Press of America, 1984). David M. Armstrong provides a helpful introduction to this issue in *Universals: An Opinionated Introduction* (Boulder, Colo.: Westview Press: 1989), 64–66. Since the difference argument requires only the principle of the indiscernibility of identicals, we may remain aloof to this issue.

not identical. Likewise, if the mind and the brain have different properties, then they are not identical.

In premise (2) the proponent of the difference argument asserts that there is a difference in properties between the mind and the brain. Premise (3) follows easily from (1) and (2). The conclusion, (4), follows from (3), given the stipulations we made above regarding the conclusion of the modal argument; that is, we are assuming that materialism requires minds to be identical to brains, and there is no third option between materialism and dualism. Unlike the modal argument, we should interpret the dualism mentioned in (4) as *substance dualism,* because the key claim of the difference argument is that our psychological states have some aspect that is not a property of any physical substance. These properties must be the property of some substance or other, so we should conclude here that there is a nonphysical substance that possesses them.[49]

The question then is whether we have good reason to believe (2), and the fate of (2) will largely rest on what supposed difference between the mind and the brain the proponent of the difference argument appeals to, and there are a number of different proposals made along these lines. For example, some philosophers argue that the intentional aspect of thought is essentially lacked by anything material.[50] As we discussed in the last chapter, human thoughts definitely have intentionality, but it does not seem that physical objects have such properties, unless they are derived from human minds. For example, when Jack is thinking of the Lincoln Memorial, there is nothing about his brain that is physically similar to or about the Lincoln Memorial. It seems then that the mind, inasmuch as it is the vehicle of beliefs, is essentially not identical to the brain. Moreover, minds have sensa-

49. See note 30 above for references to philosophers who do entertain the possibility of instances of properties without any substances.

50. In chapter 5 I present a Thomistic argument against materialism based on a certain type of intentionality. Note that, though he denies anything like an identity between the human intellect and the brain, St. Thomas rejects substance dualism. This point will be discussed extensively in chapters 7–8.

tions, and it seems that physical objects do not have such properties. When Martha eats strawberry ice cream, her mind tastes strawberry, sees pink, and feels cold, but there is nothing in her brain that tastes, looks, or feels in any such way. Moreover, Martha's tasting, seeing, and feeling are private, *first-person* states, whereas all the states of her brain are in principle public, *third-person* events. Thus, Martha's mind has properties that her brain does not have.[51] Alvin Plantinga's remarks are representative of these ways of defending premise (2):

> Presumably neither [electrons nor quarks] can think—neither can believe, doubt, want, fear, or feel pain. But then a proton composed of quarks won't be able to think either, at least by way of physical relations between its component quarks, and the same will go for an atom composed of molecules, and an organ (e.g. a brain) composed of cells. If electrons and quarks can't think, we won't find anything composed of them that can think by way of the physical interaction of its parts.[52]

There are many ways that materialists reply to these sorts of arguments, but I will leave them aside for the moment until we are prepared to discuss more sophisticated versions of materialism than the position we have sketched in contrast to dualism.[53]

A both historically influential and recently defended approach to premise (2) is given by G. W. Leibniz:

> It must be confessed, moreover, that perception and what depends on it are inexplicable by mechanical reasons, that is, by figures and motions. If we pretend that there is a machine whose structure enables it to think, feel, and have perception, one could think of it as enlarged yet preserving its same proportions, so that one could enter it as one does a mill. If we did this, we should find nothing within but parts which push upon each other; we should never see anything which would explain a perception. So it is in the simple substance, and not in the composite substance or machine, that perception must be sought.[54]

51. We will discuss these arguments from *qualia* in much greater detail in chapter 5.

52. Plantinga, *Knowledge of God,* 53.

53. Particularly helpful along these lines will be our discussion of the identity theory in section 3 of chapter 4 and emergentism in chapter 6.

54. G. W. Leibniz, *Philosophical Papers and Letter,* trans. Leroy Loemker (Dordrecht, The Netherlands: Kluwer Academic Publishers, 1969), 644.

Leibniz's point is that something composed of distinct physical parts is not the sort of thing that can have a perception (a thought or a sensation). The mind must not be something composed of physical parts, which means that it is distinct from the brain. This of course raises the question as to what it is about things composed of physical parts that excludes them from having perceptions, and we receive a clue to the answer in Leibniz's final sentence in the quotation, wherein he claims that perception must be attributed only to a *simple substance.* That is, Leibniz defends (2) by arguing that the mind has unified experiences that cannot be features of a composite physical object like a brain, and this is in fact how other early modern philosophers defend similar premises.[55] Similar arguments for (2) based on the unity of consciousness have also received a good deal of attention from cotemporary philosophers.[56] We will call this way of defending (2) the unity argument.

To understand how the unity argument is used to defend premise (2), suppose that Cormac is aware of a state of affairs, *S,* in which *the red truck making the buzzing noise is next to the shoe that smells funny.* Cormac's awareness of *S* is both multifaceted and unified in the sense that *he* is aware of colors, sounds smells, positions, kinds of things, and relations simultaneously in a way that wraps things up into a single state of affairs. The question then is whether this awareness of a unified state of affairs could be a property of a complex physical object. If such an awareness could not be a property of a complex physical object, then it follows that Cormac's mind has a property, unity of consciousness, that his brain lacks. We would then we have good reason to accept premise (2) of the difference argument.

55. See Descartes, *Meditations,* 59, and Immanuel Kant, *The Critique of Pure Reason,* trans. Norman Kemp Smith (New York: St. Martin's Press, 1965), 335.

56. William Hasker defends such an argument in various places, including *The Emergent Self,* 122–46, and "Persons and the Unity of Consciousness," in *The Waning of Materialism,* ed. Robert Koons and George Bealer, 175–90 (New York: Oxford University Press, 2010). See also, Goetz, "Modal Dualism: A Critique," in *Soul, Body, and Survival,* 89–104, and "Substance Dualism," in *In Search of the Soul,* 43–46; E. J. Lowe, *Personal Agency,* 19–23, and "Substance Dualism: A Non-Cartesian Approach," in *The Waning of Materialism,* 439–61.

Simplifying matters a good bit, we expect that each of the aspects of Cormac's singular awareness is processed by a different module or component of Cormac's brain. Suppose then in the following model that each of the regions of Cormac's brain mentioned is distinct from all of the others:

B1 is the processing of the quality *red* by a region of the brain.
B2 is the processing of the quality *buzzing* by a region of the brain.
B3 is the processing of the quality *smells funny* by a region of the brain.
B4 is the processing of the kind *truck* by a region of the brain.
B5 is the processing of the kind *shoe* by a region of the brain.
B6 is the processing of the relation *next to* by a region of the brain.

Could the set of these neurophysiological events, {*B1, B2, B3, B4, B5, B6*}, be aware of *S*? To see why the defender of the unity of consciousness argument argues for the negative, suppose I have five objects hidden in a box, that is, a spoon, a marble, a pocket knife, a book, a salamander, and

William is aware of the spoon.
Martha is aware of the marble.
Patrick is aware of the pocket knife.
Brendan is aware of the book.
Jack is aware of the salamander.

Is the set of children, {William, Martha, Patrick, Brendan, Jack}, aware of the entire contents of the box? It depends on what you mean by "aware of." On the one hand, if one means that "each thing in the box is the object of awareness of some child or other," then the set of children is aware of the box. Notice, however, that in this case *nobody* is aware of the *entire contents* of the box. On the other hand, if one means "some child is aware of the entire contents of the box," then in this case there is no such child. Moreover, one can't claim that there is some overall awareness that the set of children has, because sets or conjunctions of things don't have *awarenesses* lacked by their members. Likewise, {*B1, B2, B3, B4, B5, B6*} is not aware of *S* in this sense; even if each aspect of *S* is the object of awareness of some module of Cormac's brain or

other, there is still nothing in the brain that has awareness of *S* as a whole. To say that somehow the brain as a whole is aware of *S* as a whole would be to make the same mistake as claiming the set of children as a whole is aware of the entire contents of the box. As William Hasker puts it, "*A person's being aware of a complex fact cannot consist of parts of that person being aware of parts of the fact.* A conjunction of partial awarenesses does not add up to a total awareness."[57] It seems then that Cormac's mind has a property, unified awareness of a state of affairs, that his brain lacks. Hasker summarizes this line of reasoning as follows:

> Suppose, then, that the state [of consciousness] is broken up into bits in such a way that some part of it is represented in each of many different parts of the brain. Assuming this to be done, we have still the question: who or what is aware of the conscious state as a whole? For it is a fact that your are aware of your conscious state, at any given moment, as a unitary whole. So we have this question for the materialist: when I am aware of a complex conscious state, what physical entity is it that is aware of that state? This question, I am convinced, does not and cannot receive a plausible answer.[58]

It appears that we now have good reason to accept premise (2) of the difference argument, but a number of routes lay open to the materialist to reply. One tempting option would be to argue that we are likely to find a central processing center of the brain that has the awareness of the whole of a state of affairs. The problem with this is (a) there is no evidence as yet of such a module, and there is some good reason to doubt that is how the brain works;[59] and (b) such a module would probably be a complex object with an internal division of labor, so the very same problem will arise—we are on our way to a regress of central processing units. The regress would need to terminate in something simple, and it seems odd that simple physical particles would be

57. Hasker, *The Emergent Self,* 128. Author's emphasis.

58. Hasker, "Persons and the Unity of Consciousness," 181–83.

59. Owen Flanagan, *Consciousness Reconsidered* (Cambridge, Mass.: MIT Press, 1992), 58. Hasker discusses Flanagan's position in *The Emergent Self,* 128.

conscious.[60] Some materialists reject the notion that we possess a unified consciousness, because they deny that there are really any psychological states to be unified in the first place, while others argue that certain features of living organisms are sufficient to account for a unified consciousness.[61] I leave discussion of these replies for later chapters, because we first need to introduce more sophisticated versions of materialism than the preliminary presentation we have discussed so far.

Before we turn our attention wholeheartedly to the materialist's primary objections to dualism, we will discuss a final dualist argument, which has recently been proposed by Peter Unger.[62] Unger's argument is very complicated and quite technical, but it can be boiled down to the following line of reason. He begins by pointing out that modern science reveals that complex physical objects are not the tightly bound wholes that common sense might lead us to expect. When we consider the smallest parts of physical complexes along their peripheries, it is arbitrary as to whether, for instance, a single electron is in fact part of such a complex; along the "edges" of every macro-physical object, for example, tables, rocks, trees, there are billions of particles that have equal claim to being parts of such objects as not. Physical objects don't seem to have sharp boundaries that allow for a definite determination as to where they "begin and end." Take, for example, the black walnut tree in my back yard, call it *T.* There are billions of particles hovering around the mass of *T* that have perfectly good claims to being parts of *T,* though they also could just as easily not be parts of *T.* We will call one of these particles, *P.* Let's call the complex of black walnut tree parts with *P* "*T-plus-P*" and

60. Such a proposal is entertained by Roderick Chisholm when defending a simplicity argument. See Chisholm, "On the Simplicity of the Soul," *Philosophical Perspectives,* vol. 5: *Philosophy of Religion,* ed. James E. Tomberlin, 167–81 (Atascadero, Calif.: Ridgeview Publishing, 1991).

61. Hasker provides a very helpful consideration of empirical objections to his argument in "Persons and the Unity of Consciousness."

62. See chapter 7 of Unger, *All the Power in the World.*

the complex of black walnut tree parts without *P* "*T-minus-P.*" Notice that *T-plus-P* and *T-minus-P* are equally qualified as black walnut trees. In Unger's view, we should conclude that *T-plus-P* and *T-minus-P* are actually existing, distinct, yet overlapping, black walnut trees. Since *P* is just one of billions of particles vaguely associated with my black walnut tree, Unger would have us conclude that *where we typically take there to be a black walnut tree, there are in fact billions of overlapping trees.* Thus, Unger argues that our commonsense beliefs regarding the unity of complex physical objects is greatly mistaken; that is, where we normally think we encounter a single, unitary physical object, there is in fact a plurality of billions of overlapping objects.[63] Let's call this conclusion of Unger's the *plurality thesis.*

The following argument captures how Unger uses the plurality thesis to argue for dualism as follows.

(5) Suppose I am in a certain psychological state, *E,* for example, experiencing the taste of coffee.

(6) Suppose that materialism is true.

Therefore:

(7) My brain, *B,* is in *E.* (from 5 and 6)

Therefore:

(8) There are billions of brains in *E.* (from 7 and the plurality thesis)

(9) Our conclusion (8) is absurd. (premise)

Therefore:

(10) Supposition (6) is false. (from 6 and 9)

Therefore:

(11) Dualism is true.

Premise (5) is a noncontroversial supposition, and (6) is something we are supposing for the sake of argument, only to reject it ultimately. The inference of (7) is reasonable, given how we have construed materialism thus far in our discussion. Before encoun-

63. This is a line of argument that Unger has advanced independently of his views in the philosophy of mind. For his original presentation of the argument see "The Problem of the Many," *Midwest Studies in Philosophy* 5: 411 (1980) –67.

tering Unger's argument one might not take (8) very seriously, but it is clearly a consequence of (7) and the plurality thesis. A difference of one particle at the periphery makes no difference as to whether a brain can support a particular psychological state, so if the plurality thesis is correct, within the space roughly outlined by my skull (or skulls) there are in fact billions of brains, each of which is so constituted as to support *E*. Thus, it would seem then that, given (7), there are billions of brains within the space outlined by my skull, each of which is tasting coffee. Premise (9) asserts that (8) is absurd, and Unger's primary reason for this is that he believes it is perfectly obvious that I do not share my psychological states with billions of other distinct persons. Indeed, it would otherwise seem to be rather crowded in here![64] Since the supposition of (6) entails an absurdity, we must take it to be false as is claimed in (10), and since we have supposed throughout this chapter that our only choices are between dualism and materialism (an assumption we will latter scrutinize), it then follows in (11) that some version of dualism is true. Unger argues that this argument proves a robust form of substance dualism, because he believes he has shown that complex, physical substances simply cannot be the bearers of psychological properties. Thus, whatever has psychological properties must be a nonphysical thing.[65]

The likely aiming point for the critic of Unger's argument is the presupposition of the plurality thesis. The metaphysical issues raised by the composition of physical substances by fundamental particulars are among the most controversial in contemporary philosophy. I will take these issues up in some detail in chapter 7, so we will leave a proper assessment of Unger's argu-

64. Unger also offers interesting arguments as to why this plurality of billions of minds would be particularly problematic when considering free agency. See Unger, *All the Power in the World,* 414–18.

65. Once again, one might attempt to argue that there are some simple physical things that have psychological properties to avoid the inference to substance dualism. Though this is seemingly a counterintuitive claim it has been suggested (see note 60), and we will discuss a similar view when we encounter panpsychism in chapter 6.

ment until then. In the meantime, suffice it to say that, despite the air of strangeness it carries, Unger's argument does seem to raise some strong considerations in favor of substance dualism.[66]

Concluding Remarks about the Case for Dualism

One frequently encounters the presumption that dualism is some sort of religious superstition that certain people take based purely on blind faith or a failure to reflect on the relevant issues. I hope that the foregoing discussion of dualism has disabused you of this prejudice, as we have seen dualism defended by rigorous arguments that deserve to be taken seriously by any honest inquirer. Moreover, religious or theological motivations have played no role in these arguments, which have appealed only to philosophical reason. We have not found, however, that any of the arguments for dualism we have considered thus fair rise to the status of *demonstrative* proofs. That is, in each case we have discussed (or we will later discuss) considerations that may give us pause regarding key premises in these arguments. This is not to say, however, that these arguments fail to support dualism significantly, but only that they do not, so far, constitute absolute proof. A reasonable person may look at these arguments and simply say, "I don't buy it!" for some pretty good reasons. Nevertheless, we have likewise seen (or we will see) that many if not all of these objections can be given plausible, even if not compelling, replies on behalf of dualism. Thus, a reasonable person might also look at these arguments and come away thinking she has good reason to accept dualism. We will conclude for the moment that the dualist has a strong case to make. All is not lost, however, for the materialist, and we will consider a series of very pressing criticisms of dualism in the following chapter.

66. For a concise summary and critique of Unger's argument, see Timothy O'Connor's review of *All the Power in the World, Notre Dame Philosophical Reviews,* March 10, 2007, http://ndpr.nd.edu/news/25237-all-the-power-in-the-world/.

CHAPTER 3

THE CASE AGAINST DUALISM

THE PROBLEM OF MIND-BODY INTERACTION

In our last chapter we considered several valid (*if* their premises are true, *then* their conclusion likewise must be true) arguments for dualism, but their soundness (whether in addition to being valid the arguments have true premises) is something about which there can be legitimate disagreement. We concluded that even if these arguments do not rise to the standard of complete philosophical demonstration, it is not unreasonable for a dualist to justify her position on such grounds. Along the way we considered objections to some of the premises of the dualists' arguments that we might expect materialists to introduce, which we found to have significant force, even if not themselves decisive. Notice, however, to refute (or call into question) the argument for a position is not the same thing as to prove that such a position is false. For example, I might give you good reasons to

doubt somebody's argument in support of the existence of the Easter Bunny, but that is not to say that I have necessarily given you good reasons to conclude that there is no such being.[1] In this chapter, we will turn our attention not just to objections to arguments in support of dualism, but to attempts to prove that dualism is false (or at least that it is unreasonable to accept such a doctrine). If these arguments against dualism are successful, then we would have good reason to reject dualism, even in light of the arguments we discussed in the previous chapter.

As we shall see, the attempt to show that dualism is false (or at least not something we can reasonably believe) is the first step in the materialist's overall project. After putatively showing the absurdity of dualism, the materialist typically attempts to give an account of psychological states in entirely physical terms (or to reject the reality of psychological states entirely). That is, materialists try to show how sensations and thoughts could turn out to be material objects, or that there really are no such states, despite the appearance otherwise. If thoughts and sensations could plausibly be claimed to be material objects or merely illusions, then our certainty in the key premises of the various arguments for dualism should be seriously shaken. Successfully casting doubt on these premises, coupled with a successful criticism of dualism, would give the materialist a particularly strong case for his position. Thus, materialists employ a two-phase strategy in their rejection of dualism: (a) They point to the problem of mind-body interaction and other considerations that they suppose to give us good reasons to think dualism is false, even absurd; and (b) they offer various theories of mind that treat thoughts and sensations as either nonexistent or physical states, while attempting to show that such an account can still make sense of our commonsense notions about minds. In this chapter we will discuss in some de-

1. Remember that we are assuming that the dualist and materialist share the burden of proof, though the burden of proof in the debate over the Easter Bunny would presumably fall squarely on the pro-bunny side.

tail the prospects for the first phase of the materialist strategy. The second phase will be our concern in the following chapter.

Intelligibility

One of the most compelling facts about psychological states is their effectiveness; that is, they seem to make a difference for our behavior. Brendan winced *because of the pain* caused by scraping his knee. Will did his homework *because he wants* desert. Martha said, "Seven!" *because she thought* that was the right answer. In all of these cases, each of which seems perfectly plausible, a sensation or thought causes a behavior involving a physical motion in the person's body, most directly in his or her brain. Moreover, it is equally obvious that our bodily states have an effect on what psychological states we might be in; we feel pains as a result of injury to our bodies, and our actions (as both moral and intellectual agents) are often prompted by an encounter with some physical object. One of our primary experiences of ourselves is as agents speaking or acting because of some prior psychological state, and subsequently it is hard for us to take at all seriously even the suggestion that our minds are utterly powerless to make a difference in our physical behaviors. Jaegwon Kim makes the centrality of causal power for our understanding of mind clear in the following passage:

> Mind-to-body causation is fundamental if our mentality is to make a difference to what goes on in the world. If I want to have the slightest causal influence on anything outside me—to change a light bulb or start a war—I must move my limbs or other parts of my body; somehow, my beliefs and desires must cause the muscles in my arms and legs to contract, or cause my vocal cords to vibrate. Mental causation is fundamental to our conception of mentality, and to ourselves as agents.[2]

Epiphenomenalism is the claim that psychological states cannot affect the material world, and thereby cannot provide expla-

2. Jaegwon Kim, *Physicalism or Something Near Enough* (Princeton, N.J.: Princeton University Press, 2005), 152.

nations of our behavior. Given that it is quite obvious that our psychological states play some role in explaining our behavior, epiphenomenalism is not a desirable result for the dualist, or any philosophy of mind for that matter.[3] Epiphenomenalism is a position we should regard with great suspicion, because our common experience of ourselves as agents implies that "any theory of mind that is not able to accommodate mental causation must be considered inadequate, or at least incomplete."[4] Indeed, it is the accusation of just this failure that leads many materialists to regard dualism as a disaster.[5]

Descartes encountered a version of this objection from his contemporaries.[6] According to the reigning physics of Descartes's day, all interaction involving physical objects is really the transfer of motion; one physical object in motion sets another physical object in motion by way of collision or impact. For Descartes (and we may suppose most dualists would agree with him) the mind is essentially nonspatial, but a nonspatial thing cannot be in contact with something, let alone have a motion transferring collision. Thus, it is a great mystery as to how a nonphysical mind and a physical object could interact, because the mind can neither set something else in motion nor be set into motion itself. Descartes struggled with this very problem, and even though he went so far as to suggest the point of contact between the mind and the brain, he never really explained how such a contact is possible. In the seventeenth century, Descartes's followers, such as Nicole Mal-

3. Some dualists, especially property dualists, have defended epiphenomenalism. For a famous example, see Thomas Huxley, "On the Hypothesis that Animals Are Automata and Its History," http://alepho.clarku.edu/huxley/CE1/AnAuto.html. D. M. Armstrong offers a detailed critique of Huxley's epiphenomenalism in ch. 4 of *The Mind-Body Problem: An Opinionated Introduction* (Boulder, Colo.: Westview Press, 1999). Armstrong has also appended Huxley's essay to this volume. Kim himself now defends a version of epiphenominalist property dualism about certain types of psychological states (qualia) in *Physicalism or Something Near Enough*, 170–73.

4. Kim, *Physicalism or Something Near Enough*, 152.

5. John Searle, *Mind: A Brief Introduction* (New York: Oxford University Press, 2004), 8.

6. See Descartes's reply to Princess Elizabeth in *Philosophical Essays and Correspondence*, ed. Roger Ariew (Indianapolis, Ind.: Hackett Publishing, 2000), 213–16.

ebranche, and other philosophers influenced by Descartes, such as G. W. Leibniz, attempted to avoid this problem by coming up with various noninteractive versions of the relationship between mind and body.[7] The problem is that each of these theories either is plagued with deep problems of its own or is just plain difficult to square with common sense. For example, Malebranche's *occasionalism* accounts for the appearance of mind-body interaction by claiming that God simultaneously causes corresponding mental and physical events, whereas Leibniz argues that the mind and the body are noninteracting substances following preprogrammed series of changes that parallel each other so closely so as to give the appearance of interaction. Both of these theories would "solve" the problem of mind-body interaction by denying that any such interaction occurs, despite appearances otherwise. Even though these theories are embedded in the broader metaphysical views of some of the greatest early modern philosophers, they are much too extravagant for us to take terribly seriously.[8]

Luckily for the substance dualist the sort of impact physics that troubled Descartes's theory of mind is no longer scientifically credible, so our inability to account for contact between mind and body need not be a deterrent to dualism. For example, Newtonian physics posits gravity as a force that acts between bodies that may be separated by great distances, and modern quantum physics allows for "entangled" particles that interact, in some sense, without any direct physical contact.[9] Some materialists, however, insist that there is still a problem for the dualist in this

7. Leibniz can be said to be a dualist only in a very broad sense, because he believes that bodies themselves are ultimately composed of immaterial substances. See Leibniz, *Monadology,* in *Philosophical Papers and Letters,* 643–54; and Nicolas Malebranche, *The Search after Truth,* trans. Thomas Lemmon (New York: Cambridge University Press, 1997).

8. For a useful survey of the history of the problem of mind-body interaction in the seventeenth century, see Karl Popper and John Eccles, *The Self and Its Brain: An Argument for Interaction* (New York: Routledge, 1977), 176–88.

9. See Lowe, *Personal Agency,* 58-62, for a discussion of why Descartes's problem of mind-body interaction became irrelevant with advances in physics.

vicinity. For instance, David Armstrong believes that mind-body interaction is still a serious problem for substance dualism:

> But it remains true that we still have no clear model for the way that a spiritual happening gives rise to a brain happening, and vice versa. Descartes, in reply, can protest only that the interaction occurs. But that is no reply if what we are trying to assess is the plausibility of his theory of that interaction. There cannot be much doubt about the occurrence of mind-body interaction. But there can be doubt about whether this is the interaction of two totally disparate things. A mind-body theory that makes interaction easy and intelligible (as a materialist theory seems to do) is to that extent preferable.[10]

In other words, even given the changes in our understanding of physical interaction since Descartes's day, mind-brain interaction remains *strange.* On the one hand we have an entirely material substance, and on the other hand we have an entirely mental substance. These two entities are supposed to have nothing in common: for example, the one is spatial, whereas the other is nonspatial; the one has intentionality, whereas the other has no intentions. It just seems quite odd, say materialists such as Armstrong, that two things that are so very different (possibly with nothing significantly in common) could affect each other. There is no intrinsic property that the one has that the other could gain. Although few philosophers doubt that there is mind-body interaction, if dualism is true, we don't have a clue as to how this happens. We cannot produce a model or image to explain how a nonphysical substance or nonphysical properties would interact with a physical system like a brain, and it thereby appears that mind-body interaction is an interminable mystery as long as we adopt dualism. Thus, whereas the dualist would need to leave mind-body interaction a mystery, the materialist, because she thinks mind-brain interaction is just one more case of physical causality, can account for it with the same principles of causality that operate in interactions between ordinary physical objects.

10. Armstrong, *The Mind-Body Problem,* 20.

There is no doubt that mind-brain interaction is strange, but some dualists are quite willing to concede this point. For instance, Sandra Menssen and Thomas Sullivan have recently admitted that "we have no idea how mind can move matter," but at the same time they wonder "how big a problem is this?" because "in the end we have no idea how *matter* moves matter."[11] In other words, Menssen and Sullivan grant that we don't have a model as to how a nonphysical mind might move a physical body (the brain), but they deny Armstrong's claim that we are any better acquainted with the workings of physical causation.

In the eighteenth century, David Hume argued that when we claim to *see* physical causation, we don't literally see the causation, but rather we see a temporal succession of events that we *infer* to be related as cause and effect based on prior experience. For example, when we see a batter cause a ball to fly into deep left field, what we directly witness is a succession of events, that is, a ball flying into the outfield after the collision with the bat, but we don't literally witness the *causing*. Hume doubts whether any sense can be made of causality at all in traditional terms, or at least whether we can ever be rationally certain we have witnessed a case of one physical event causing another.[12] We need not go that far, but there is a kernel of truth in Hume's position: How one thing can cause a change in another is far from clear based on direct empirical inspection alone. More than two centuries of philosophical reflection since Hume made this case have done little to solve the problem of causation.[13] I claim, contrary to Hume's own skeptical conclusion, that we clearly know causation when we see it, but (and here is where Hume gets something

11. Menssen and Sullivan, *The Agnostic Inquirer,* 108. Authors' emphasis.

12. Hume's clearest statement of his argument for causal skepticism can be found in sec. 4 of *An Enquiry Concerning Human Understanding* (Indianapolis, Ind.: Hackett Publishing, 1977).

13. A. J. Freddoso provides an excellent discussion of the problems for modern accounts of causation after Hume in his introduction to Fancisco Suarez, S.J., *On Creation, Conservation, and Concurrence: Metaphysical Disputations 20–22,* trans. A. J. Freddoso (South Bend, Ind.: St. Augustine's Press, 2002), lix–lxxiii.

right) we are unable to give any sort of account of *how* it works when all we have to work with are the conceptual tools acceptable to the modern mechanist.[14] As Menssen and Sullivan put it, "we do not actually experience physical causality in any deep way. We see that matter moves matter in some sense, but we do not see why or how it happens."[15] It may be that causality just is the bedrock for us; there is nothing more fundamental by which we can explain causality, because we need the notion of causality for all of our explanations.

Thus, it is no strike against dualist versions of mind-brain causation that it is opaque to our understanding, for that is the case for causality in general under the assumption of mechanism.[16] If the materialist insists that we must deny the dualist's version of mind-brain causality because it is "unintelligible," then consistency demands that she likewise deny physical causation for the same reason. That is, the materialist must either concede that the unintelligibility of mind-brain interaction poses no problem or embrace a wholesale Humean skepticism regarding causality. The latter option is no help to the materialist, because if we are skeptical about causation in general, then the mind-brain interaction problem is just a special case of a general problem of causation, and the materialist, in any event, does not have a more reasonable replacement theory.[17]

The Law of the Conservation of Energy

Sometimes the worry over mind-brain interaction takes the form of concern not over the intelligibility of mental causation,

14. I emphasize that I am not arguing that no model of causation can be given at all, but only that under mechanistic assumptions no such model can be constructed. For an account of such a model of causation see Freddoso's essay I cite in the previous note and chapter 7 of this book.

15. Menssen and Sullivan, *The Agnostic Inquirer,* 108.

16. I discuss the difficulty of giving a mechanistic account of causation in chapter 7 below.

17. Keith Yandell provides a similar reply to this particular problem of mind-brain interaction in "A Defense of Dualism," 551–52.

but over whether such interaction actually contradicts a fundamental law of physics, namely the law of the conservation of energy.[18] The conservation law is no small matter for physics. Indeed, a great deal of modern physics would be seriously hampered without the operational assumption of the constant state of energy in the universe. Given the impeccable experimental verification of our standard physical theories, it is not a good mark for any position to run afoul of the fundamental principles of physics. The naturalist philosopher John Searle sees the apparent tension between dualism and the conservation law as a decisive argument against the former.

> It seems impossible to make substance dualism consistent with modern physics.
>
> Physics says that the amount of matter/energy in the universe is constant; but substance dualism seems to imply that there is another kind of energy, mental energy or spiritual energy, that is not fixed by physics. So if substance dualism is true then it seems that one of the most fundamental laws of physics, the law of conservation, must be false.[19]

The point is that modern physics generally operates under the assumption that the quantity of energy in the universe remains constant. If a nonphysical mind can affect the brain, then it is adding motion to a physical system, which would seem to increase the total quantity of energy in the universe. Thus, modern physics shows that mind-brain interaction, as conceived by substance dualists, cannot occur.

There are three typical ways dualists reply to the objection from the conservation law: (a) Some dualists have attempted to deal with this problem by arguing that it is possible that "the energy transfers due to consciousness involve offsetting losses and gains," and one way this could happen is by "the system requiring energy to produce the conscious state and the conscious state

18. Hereafter we will simply call the law of the conservation of energy "the conservation law."

19. Searle, *Mind: A Brief Introduction*, 29–30.

infusing energy back into the system."[20] On this model the loss of energy on the side of the brain would be compensated by the mind paying back an equal amount of energy through its interaction with the brain. This proposal seems plausible, though I will leave it to those better able to assess the relevant scientific issues to pursue it further.

(b) The objection from the conservation law poses a much greater challenge than the objection from the supposed unintelligibility of mind-brain interaction, because it purports to show that substance dualism contradicts a truth we cannot easily give up. In reply, the dualist may first point out that even some physicists, in light of the radical discoveries of the twentieth century and the continued problems of making sense of living and conscious beings in a physical world, have concluded that some of the fundamental laws of physics themselves may need to be revised.[21] For example, the dualist Nobel laureate neuroscientist John Eccles cites Eugene Wigner as a prominent physicist who believes that "the laws of nature will have to be changed, not only reinterpreted" in order to make sense of psychological phenomena.[22] In the last century physicists made startling discoveries regarding the behavior and nature of very small, subatomic particles, for example, electrons. They found that these entities do not behave at all in the deterministic manner of macro-physical objects. In some cases, it even seems that their behavior may depend on conscious perception by observers, and it appears that these particles (quanta of energy) can come in and out of existence. It is then unclear if the conservation of energy really is all

20. Menssen and Sullivan, *The Agnostic Inquirer,* 112. Menssen and Sullivan cite Popper and Eccles, *The Self and Its Brain,* 541–547, as an example of such an interactionist proposal.

21. E. J. Lowe notes a tendency among some contemporary thinkers to disregard conservation laws as serious concern for mind-body interaction in *Personal Agency,* 41–43.

22. Quoted in Popper and Eccles, *The Self and Its Brain,* 544. This passage was brought to my attention by Menssen and Sullivan, *The Agnostic Inquirer,* 111. My treatment of the objection from the conservation law owes a great deal to Menssen and Sullivan's treatment of this issue on pp. 108–13 of the *Agnostic Inquirer,* though I do not follow their discussion closely at all points.

that fundamental.[23] Wigner argues that we now must conclude that "through the creation of quantum mechanics, the concept of consciousness came to the fore again: It was not possible to formulate the laws of quantum mechanics in a fully consistent way without reference to consciousness."[24] In other words, in light of the results of quantum mechanics, we must adjust fundamental physical laws to account for the apparent irreducibility of consciousness, and there are many other very prominent physicists who also revise the conservation law for various reasons.[25] Rather than denying either the efficacy or reality of nonphysical minds, scientists such as Eccles and Wigner would sooner revise the conservation laws. Eccles himself produces a model of how interaction might work using quantum indeterminacies.

(c) Entrenched controversy rages regarding the proper interpretation of quantum mechanics, and there are those who believe the strange behavior of very small particles need not call into question the fundamental laws governing larger physical objects. This is largely a scientific controversy, and the philosopher does well not to predicate her conclusions on a partisan position in a debate among physicists. Thus, the substance dualist should not be too quick to jettison the conservation law, even on the advice of great scientists such as Eccles and Wigner. Luckily for the dualist, there is a middle position between outright rejection of the conservation law and an absolute, exceptionless interpretation. There are those who interpret physical laws statistically; that is, physical laws do not tell us what absolutely must happen given prior conditions, but rather they tell us what is highly prob-

23. We have waded into deep scientific waters here. Bruce Rosenblum and Fred Kuttner provide a good introduction to quantum mechanics and its implications for consciousness in *Quantum Enigma: Physics Encounters Consciousness* (New York: Oxford University Press, 2006).

24. Eugene Wigner, "Remarks on the Mind-Body Question," in *Philosophy of Mind: A Guide and an Anthology,* ed. John Heil (New York: Oxford University Press, 2004), 867.

25. See Helge Kragh, *Cosmology and Controversy: The Historical Development of Two Theories of the Universe* (Princeton, N.J.: Princeton University Press, 1996). Menssen and Sullivan also cite this source to support a looser interpretation of the conservation law.

able to happen. According to a statistical interpretation, the conservation law would be consistent with very small, probably immeasurable, losses and gains of energy. Presumably, this is what the dualist has in mind in cases of mind-brain interaction, so on a statistical interpretation, the dualist would not need to give up the conservation law. As Menssen and Sullivan put it, "a statistical interpretation seems to fit at least as well. This is one reason physicists have felt free to see 'exceptions' to the law. If the law is only statistical, then it is possible that the causal impact of consciousness is immeasurable."[26]

There are then a couple of reasonable, even if controversial, options for the dualist to defend against the objection based on the conservation law, so it is premature for a materialist to conclude that dualism is defeated by this objection.[27] The dualist Karl Popper, who himself is not one to question the validity of physics, concludes that the idea that sensations and thoughts are physical is more absurd "than the assumption of some slight and probably immeasurable violation of the first law of thermodynamics."[28]

The Causal Closure of the Physical

Though supposed unintelligibility and the conservation law do not pose decisive problems for mental-physical causation, the dualist is not out of the woods yet with respect to interaction. Jaegwon Kim points out that most materialists reject dualism because "they accept the causal closure of the physical not only as a fundamental metaphysical doctrine, but as an indispensable

26. Menssen and Sullivan, *The Agnostic Inquirer,* 112.

27. For a very good overview discussion of why modern physics does not seem to rule out interactive dualism, see Robin Collins, "The Energy of the Soul," in *The Soul Hypothesis: Investigations into the Existence of the Soul,* ed. Mark C. Baker and Stewart Goetz, 123–33 (London: Continuum, 2011). Though his concern is primarily with whether conservation laws exclude the possibility of God's interaction with the physical world, Alvin Plantinga provides a very helpful discussion of these issues in *Where the Conflict Really Lies: Science, Religion, and Naturalism* (New York: Oxford University Press, 2011), 65–90.

28. Popper, *The Self and Its Brain,* 544.

methodological presupposition of the physical sciences."[29] That is, an assumption of the natural sciences is that *those events that have causes are physically caused.*[30] Kim, speaking for a great many philosophers and scientists, argues that the causal closure of the physical is a cornerstone principle of natural science. The scientist looks for physical causes, and the very endeavor of doing so presupposes that there are such causes. If science were open to nonphysical causes, science would then be open to explanations that cannot be specified in scientific terms. Thus, the intelligibility of the scientific method presumes that the scientist operates under the assumption of causal closure.

If one takes causal closure seriously, then it would seem that there is no work left to be done by nonphysical causes in explaining phenomena that are susceptible to scientific explanation. Since physical causes are the object of scientific investigation, it follows that where we have successful scientific theories or explanations (or where we think one is likely), there is no need for nonphysical entities or events to act as causes. We don't entertain nonphysical explanations for the motion of inanimate bodies, diseases, the growth of living beings, and so on, because we have scientifically based, physical explanations available. Neuroscience likewise operates under the assumption of causal closure. As Kim puts it:

> Brain scientists will not look outside the physical domain for explanation of neural phenomena. They are not likely to think that it will be scientifically productive to look for nonphysical, immaterial forces to explain neural events. We expect the physical world to be causally self-contained and explanatorily self-sufficient. That is, we suppose that if a neural event—or more broadly, a physical event—has a cause, or an explanation, then it must have a physical cause or explanation. This is the principle of causal/explanatory closure of the physical domain.[31]

29. Jaegwon Kim, *The Philosophy of Mind* (Boulder, Colo.: Westview Publishing, 1998), 148.

30. Kim, *Physicalism or Something Near Enough,* 43.

31. Ibid., 154.

If we have a physical explanation available for the happenings in the human brain associated with our psychological lives, then it would seem there is really nothing for supposed immaterial entities to do, or, at least for scientific purposes, the available physical explanation makes any supposed nonphysical cause superfluous.

Kim argues that we have just such an explanation:

> You want to raise your arm, and your arm goes up. Presumably, nerve impulses reaching appropriate muscles in your arm made those muscles contract, and that's how the arm went up. And these nerve signals presumably originated in the activation of certain neurons in your brain. What caused these neurons to fire? We now have a quite detailed understanding of the process that leads to the firing of a neuron, in terms of complex electrochemical processes involving ions in the fluid inside and outside a neuron, differences in voltage across cell membranes, and so forth. All in all we seem to have a pretty good picture of physics, chemistry, and biology. If the immaterial mind is going to cause a neuron to emit a signal (or prevent it from doing so), it must somehow intervene in these electrochemical processes. Surely the working neuroscientist does not believe that to have a complete understanding of these complex processes she needs to include in her account the workings of immaterial souls and how they influence the molecular processes involved.[32]

Kim's point is that the neuroscientist knows, in ever-increasing detail, the physical causes of neurophysiological events. These physical causes seem to be *sufficient* for their physical effects, meaning that nothing else is needed to bring about the effect. Suppose we know that when Jack desires a cracker within his reach at a certain time, T_1, he is in a certain neurophysiological state that reliably predicts (to some degree at least) a subsequent brain state at T_2 associated with his grasping the cracker. A neuroscientist with such knowledge can give an account of this process within Jack's brain that makes reference only to physical causes, that is, an entirely scientific explanation making no use of nonphysical causes. If the physical is causally closed, at least where scientific explanations are available, then it seems that

32. Kim, *Philosophy of Mind,* 131–32.

nonphysical events and substances have no work to do in explaining our behavior. As Kim puts it: "Even if the idea of a soul's influencing the motion of a molecule … were coherent, the postulation of such a causal agent would seem neither necessary nor helpful in understanding why and how our limbs move."[33] If we accept both the necessity of the causal closure of the physical for scientific explanation and the proposal of scientific explanations of brain states offered by neuroscientists, it would seem that the best the dualist can hope for is epiphenomenalism, which of course is not a good fate.[34]

For the sake of clarity, let's take a moment to put this objection in the form of an argument:

(1) Any brain state, *B,* is caused by some prior physical occurrence, *P.* (premise)

(2) If a physical state (including brain states) is caused by a physical occurrence, then that state does not have a nonphysical cause. (premise)

Therefore:

(3) *B,* which could be *any* brain state, does not have a nonphysical cause. (from 1 and 2)

The conclusion, (3), clearly follows from (1) and (2), so the only question is whether we have good reason to believe these two premises. Someone who accepts causal closure will assume, as we claim in (1), that all brain states have physical causes in prior physical occurrences. To doubt (1) would be to hold out hope that neuroscientists will discover a brain event for which there is no prior physical cause, which seems highly unlikely. In any event, it is unwise for the philosopher to place his bets on what will or will not be discovered scientifically. We do well for now at least to grant (1). Premise (2) also seems quite reasonable. If

33. Ibid.

34. I will take Kim's claim that causal closure is essential for neuroscience at face value for the moment. See Mario Beauregard and Denyse O'Leary, *The Spiritual Brain: A Neuroscientist's Case for the Existence of the Soul* (New York: HarperOne, 2008), for an example of a neuroscientist who doubts this methodological assumption.

we have a physical explanation for a certain state or event, then it would seem that nonphysical factors would do no work, or at least it is unclear as to why we would think we need to posit an additional nonphysical explanation. If the prior physical state is *sufficient* to bring about the event we are hoping to explain, we have no reason to appeal to nonphysical causes. The materialist is right to ask "Causes of what?" regarding supposed nonphysical states or substances, when causally sufficient physical explanations for neurological events are available.[35] It seems then that we have a pretty strong case for (3), which itself entails that the best the dualist can hope for is epiphenomenalism, because nonphysical events then have no role in explaining the neurophysiological events and processes that lead to our actions.

The issue of the causal closure of the physical world is hotly debated among contemporary philosophers of mind.[36] For now, we will consider one recent attempt to reply on behalf of the dualist offered by Stewart Goetz and Charles Taliaferro. Goetz and Taliaferro do not dispute that (1) and (2) imply (3), nor do they say anything that calls (2) into doubt. Their argument, however, is focused on something like (1). Goetz and Taliaferro do not point to any neurophysiological states that can be shown to lack a physical cause *under laboratory conditions,* that is, when neuroscientists are studying them, but they doubt that this fact is enough to support (1). Goetz and Taliaferro grant that scientific work does require "the assumption of causal closure of the areas of brains [the neuroscientist is] studying during his experiments."[37]

35. In order to simplify things for the novice reader, I have assumed throughout this discussion of causal closure that the notion of *causal overdetermination,* i.e., a single effect can have two distinct but sufficient causes of the same type, is untenable. There are, however, prominent philosophers who have defended such a view. Trenton Merricks criticizes overdetermination in *Objects and Person* (New York: Oxford University Press, 2001). For a sympathetic discussion of overdetermination, see Ted Sider, "What's So Bad about Overdetermination?" *Philosophy and Phenomenological Research* 67, no. 3 (2003): 719–26. See also Kim's rejection of overdetermination in *Physicalism or Something Near Enough,* 46–52.

36. William Hasker offers a good overview of the debate along with his own sustained argument against causal closure in *The Emergent Self,* 58–80.

37. Goetz and Taliaferro, *Naturalism,* 35.

For example, if a neurophysiological state immediately follows the artificial stimulation of somebody's brain by a neuroscientist, it is best to assume that the physical cause, that is, the artificial stimulation, is the only cause of the event. Goetz and Taliaferro argue that the concession of causal closure for experimental purposes does not entail that the scientist "must be committed as a scientist to the universal explanatory exclusion of mental events that on certain occasions cause the occurrence of events in the physical world."[38] Rather:

> All that a neuroscientist as a scientist must assume is that during his experiments mental events … are not causally producing the relevant events in the microphysical entities in the delimited areas of the brain he is studying. If a neuroscientist makes the universal assumption that in any context events in microphysical entities can only have other physical events as causes and can never be causally explained by mental events and their purposes, then he does so not as a scientist but as a strict naturalist.[39]

Goetz and Taliaferro's point is that science works by discovering certain conditional propositions, for example, *if P occurs under certain conditions, then Q occurs.* A neuroscientist might discover that "if region *X* of the brain is stimulated under conditions *C,* then region *Z* activates while the subject claims to see blue" or "if region *X* is observed by an MRI as being activated, then region *Z* will immediately be observed by an MRI as being activated while the subject claims to see blue." Both of these conditional statements may be interpreted as citing sufficient physical causes for the occurrence of neurophysiological states and accompanying psychological states, but that is not to say that such a stimulus has been discovered to be a *necessary condition* for the subsequent neurophysiological state and/or the accompanying psychological state. In other words, the scientific evidence is perfectly consistent with those neurophysiological states having other causes, even nonphysical causes, under other conditions. Even though

38. Ibid., 35.
39. Ibid.

we observe the neurophysiological state as having a physical cause under experimental conditions, it remains possible that a neurophysiological state of the same type may have a nonphysical cause under other conditions.[40]

Thus, Goetz and Taliaferro would have us revise premise (1) as follows:

(1*) If a subject *S* is given physical stimulus, *P,* then brain state *B* occurs.

They grant that brain events can be caused by physical stimuli, but (1*) is too weak to support (3). That is, (1*) at most implies only that *P* is a cause of *B;* it does not show that *P* is the only possible cause of *B.*[41] Nothing in the discovery of (1*) precludes that *B* might have a different cause from *P,* even a nonphysical one, in different situations. The fact that lightning causes forest fires does not rule out the possibility of forest fires being caused by sparks from a poorly managed campfire. We know, given our discussion above and as Goetz and Taliaferro point out, that mind-brain interaction may well be physically possible in light of the results of recent scientific developments. Thus, it would seem that the scientific presupposition of causal closure is not a serious threat to substance dualism.

Goetz and Taliaferro provide the dualist with a powerful line of defense, but I doubt that it is ultimately enough to save the day alone. Certainly, the presupposition of causal closure for scientific practice is an important part of the causal closure case against dualism, but that is not the main point. The scientist does presume causal closure for experimental purposes, and Goetz and Taliaferro are correct to point out that this presumption need not be universal; all the scientist need assume is that sufficient physical causes are available in the experimental context, which is consistent with nonphysical causes operating in other domains. The

40. See ibid., 36–38.

41. I'm assuming for the sake of simplicity that the "If … then …" statement in (1*) is sufficient for a causal relationship, though in general this is not a safe assumption.

important point, however, is that the neuroscientist *does* discover physical causes, and where physical causes are operative, we don't need nonphysical causes. Suppose that *B1* is the neurophysiological state associated with William's feeling thirsty and *B2* is the neurophysiological state that initiates the physical process of his getting a glass of water. Suppose that a neuroscientist discovers, using electromagnetic brain imaging techniques, a causal link between *B1* and *B2,* such that *B1* seems to be sufficient for *B2.* Note that the neuroscientist discovers the *B1*-*B2* causal relation not by artificially stimulating the brain under experimental conditions as in the example we discussed above, but by observing via MRI technology exactly what William's nervous system does when he is experiencing thirst. These observations reveal what appears to be a physical cause for William's getting a glass of water, namely *B1.* It seems that William's feeling thirsty has no explanatory role, assuming that we can also find a physical cause for *B1.* Thus, William's sensation of thirst plays no explanatory role.

In this case Goetz and Taliaferro's reply will not help the dualist. In the prior case, they could argue that just because we discover one sufficient cause of a brain event, which happens to be a physical cause, that does not preclude that there are ever nonphysical causes of the same type of brain event. The worry in the case of what we are supposing to discover about William with the MRI is that his sensation of thirst seems to play no role in causing his behavior, and there is no reason to believe that his brain behaves at all differently when it is not under a MRI. Using the MRI, we would expect to find a physical cause of William's going to get a glass of water, so there is no reason to suppose that his sensation does any explanatory work. The problem for the dualist is not that the scientist can get by without universal causal closure, but that scientists, including neuroscientists, discover what seem to be sufficient physical causes. Where there are physical causes we don't need nonphysical causes, and it is too far-fetched to conclude that brains operate differently in different contexts.

The simplest explanation seems to be that since we can discover physical causes in experimental conditions, physical causes are operative universally, and therefore the dualist has a problem with causal closure even in this limited variety.[42]

A contemporary philosopher, E. J. Lowe, has recently offered another way of replying to the causal closure argument on behalf of the dualist.[43] Lowe's argument is a highly sophisticated affair, which really deserves to be discussed in great detail. Given the limitations of our discussion, I will merely sketch Lowe's position in order to offer some idea as to how contemporary dualists might give a plausible answer to the causal closure argument. Suppose we have a complete physical account of why Patrick's hand rose at a particular trajectory at a certain time, beginning with neurophysiological events in his brain and ending in the movement of his arm. For example, MRI observations show that this process involves the neurophysiological events that typically lead to movements similar to the way Patrick moved his arm. It seems that we have a complete physical explanation of the movement of Patrick's arm, and there is therefore (per the causal closure argument) no work left for psychological states. Lowe, however, points out that nothing in this explanation tells us why Patrick's hand-raising was more than a merely coincidental confluence of events. While chains of physical events originating in Patrick's brain may explain why his arm rose in some particular trajectory rather than any of a number of other trajectories, no such story explains the fact that Patrick *intentionally* raised his arm. Chains of physical causes explain the particular course of

42. I am not disagreeing at all with Goetz and Taliaferro's reply to the causal closure argument based on examples of neuroscientists' attributing physical causes based on artificial stimulations of the brain. My concern is only that their reply does not do enough to displace examples in which one might observe the causes operative in the brain as it functions without artificial stimulation.

43. The following discussion is an extreme simplification of a very complicated and technical line of reasoning. Readers with less philosophical experience may find the remainder of this section very difficult to follow. I include it only to show that the dualist, though challenged by the causal closure argument, is not without resources to reply to this criticism. In short, dualism is very much a live option, even if problematic.

arm-raising that occurred, but thoughts, desires, beliefs, and so on explain the intentional act of arm-raising. As Lowe puts it, "mental causation is intentional causation—it is causation of the intended effect of a certain kind. Bodily causation is not like this. All physical causation is 'blind,' in the sense that physical causes are not 'directed towards' their effects in the way mental acts are."[44] No series of physical causes is sufficient to explain Patrick's hand-raising as an intentional act, so unless we are to claim either that there are no intentional acts or that intentional acts are fundamentally inexplicable (and neither of those options is remotely attractive), then it seems we must conclude that a psychological state (Patrick's intention) plays a role in causing the activity. In short, physical causes may be sufficient to explain the movement of Patrick's arm along a certain trajectory, but they are insufficient to explain the movement of Patrick's arm along a certain trajectory *as part of an intentional act.* The latter is the explanatory domain of psychological states, which the dualist supposes to be a type of nonphysical causation.[45]

Moreover, consider a situation in which a villain has resolved to blow up one of two buildings, but he cannot decide which. So as not to be paralyzed by indecision, the villain rigs both targets with explosives, but sets up his remote detonating device such that after he throws the switch an internal randomizing program will select one of the two buildings to be destroyed. Suppose the mechanism randomly selects building A over building B. On the one hand, the specific act of destruction (the *token* of destruction) was determined by the randomizing mechanism; that is, building A was destroyed instead of building B, because of the randomizer. On the other hand, the fact that there was any destruction at all (the *type* of activity that occurred) is explained by the plotting of the villain. In this case the villain is responsible

44. Lowe, *Personal Agency,* 110.

45. Note that I do not take intentional causation as Lowe construes it to be a case of overdetermination, because on this view the physical causes are necessary, but insufficient, conditions for the intentional act to occur.

not because he determined which particular destructive event occurred (that was a result of chance), but because he caused a state of affairs to obtain such that there would be destruction as such. Notice that neither of the two possible outcomes of the randomizing device is itself a necessary condition for the fact that there was destruction at all, because the villain has set the process up so that whatever the randomizer outputs will lead to a destroyed building. In short, there are cases requiring separate explanations for the token and the type of a certain occurrence.

With this example in mind, suppose that the exact sequence of physical events terminating in the raising of Patrick's arm, call it *P,* had been altered in terms of the activity of just one neuron such that *P** occurred instead. At most, *P** would cause a slightly different trajectory of hand-raising, as would any other similar alteration of *P.* Thus, no particular neurophysiological event is necessary for the hand-raising as such, because for any such event there is another slightly different neurophysiological event that would have brought about another token of hand-raising in that situation. If Patrick had not intended to raise his arm, then his hand-raising would not have occurred at all, so his intention is necessary for his hand-raising. Lowe uses these points to argue that, although physical causes might provide sufficient conditions for a *particular course of physical movement* independently of any supposed nonphysical causes, they are sometimes neither necessary nor alone sufficient for a certain type of activity to take place.[46] Even if we have a complete physical story as to why Patrick raised his arm *in a particular manner,* it still might be the case that *the fact that he raised his arm at all* has his intention (what the dualist supposes to be a nonphysical cause) as a necessary condition; that is, if Patrick had not intended to do so, then he wouldn't have raised his arm at all. It is just that sort of explanation that a psychological state might provide.

46. Lowe lays out this case in detail throughout chapter 5 of *Personal Agency.*

Of course, if mental states are going to be explanatory, they must at some point interact with the brain, so we must deny premise (1) of the causal closure argument as I have construed it above; there must be some neurophysiological events for which there is no sufficient physical cause. Lowe does not claim, however, that we will ever observe a neurological event that is not preceded by some physical cause. In some cases (and these may not be empirically detectable) those physical causes are, though necessary, insufficient for their effects, because there are simultaneously operating nonphysical causes also in play. As we have discussed above, there are good scientific reasons to doubt that all physical events are exhaustively determined by prior physical causes, so it is plausible to suppose that there are "gaps" in the web of physical causes that provide room for mental causes to occur.[47] "In short, mental causation would be empirically undetectable, and therefore irrelevant to the practice of neuroscience."[48] Thus, we don't need the assumption of *universal* causal closure to make sense of the practice of neuroscience, though a local assumption of causal closure is certainly appropriate for any scientific investigation.

Suffice it to say that the problems stemming from mind-brain interaction are troubling, but not insurmountable, for the dualist, and I have sketched some workable avenues for dualists to reply to the causal closure argument. Thus, the reports of the demise of dualism have been greatly overstated, especially if it is supposed that the coup de grâce has been administered by the problem of mind-body interaction. In the next section we will discuss some other, less metaphysically technical, objections to dualism.

47. Eric Reitan provides a very accessible case for the possibility of nonphysical influence in a physical universe wherein "chance gaps" occur in *Is God a Delusion? A Reply to Religion's Cultured Despisers* (Malden, Mass.: Wiley-Blackwell, 2009), 88–93. See also Plantinga, *Where the Conflict Lies,* 91–129.

48. Lowe, *Personal Agency,* 74–75.

Psycho-Physical Dependence, Conscious Animals, and the Unity of Human Persons

Thus far we have been considering objections to both substance and property dualism stemming from supposed problems of mental-physical causation. In this section we will consider several less technical concerns that are faced mainly by the substance dualist in particular. It is old news to common sense that our psychological lives are highly dependent on our brains. We all know that a bit too much beer is likely to impinge on our powers of judgment, a bump on the head might permanently or temporarily cause a loss of consciousness, many personality traits are determined by chemical balances in the brain, and psychotropic drugs can be used to alter even long-held psychological traits. These are just to name a few of the most commonly known examples. We might also add some interesting cases drawn from the sciences. Certain very serious forms of epilepsy are treated by commissuratomy, a procedure wherein the two hemispheres of the patient's brain are surgically divided. The result is often the alleviation of epilepsy, but also a sort of split consciousness; it might seem that the commissuratomy patient comes to have two minds following the division of her brain. People have also been found to suffer brain damage that leaves their visual powers intact, but denies them the ability to recognize specific objects or causes them to confuse objects. One of the most famous relevant cases is that of Phineas Gage, a nineteenth-century workman who had a tamping iron driven completely through his head. Gage survived, but his personality was permanently altered for the worse.[49]

The point of all of these examples is that psychological states are indubitably dependent on the brain, and that raises a serious

49. For a number of fascinating case studies that demonstrate psycho-physical dependence in various ways see Oliver Sachs, *The Man Who Mistook His Wife for a Hat: And Other Clinical Tales* (New York: Touchstone, 1998). Thanks to Carol Kennedy, who pointed out to me that Gage's injury was brought on by a tamping iron, not a railroad spike.

concern for the dualist. As John Searle puts it: "According to substance dualism our brains and bodies are not really conscious. Your body is just an unconscious machine like your car or your television set."[50] That is, for the substance dualist, though my body doesn't have psychological states, I have them, and it seems possible that I exist while having such states in possible worlds in which my body does not exist. The problem is that "given what we know about how the world works, it is hard to take [substance dualism] seriously as a scientific hypothesis. We know that in humans consciousness cannot exist at all without certain sorts of physical processes going on in the brain."[51] In other words, according to the substance dualist, we would not expect that having certain physical structures would be either necessary or sufficient for consciousness, but that is exactly the opposite of what both science and common sense reveal.[52] The examples we discussed above seem to show that having certain physical structures in place is at least a necessary condition for having sensations and thoughts, which is not what we would expect given substance dualism.[53]

Substance dualists will often argue that while the mind is independent of the brain, the brain is its only contact with the "outside" (physical) world. The brain provides the mind with the sensory information it needs to form sensations and thoughts, and it is only through the brain that the mind can express what is going on internally. If the brain is damaged, or otherwise altered, inaccurate information will be presented to the mind and it will be unable to express its workings physically. Thus, it is entirely possible that minds are untouched by physical circumstances, despite the appearance of being altered by changes in the brain.

50. Searle, *Mind: A Brief Introduction,* 30.

51. Ibid.

52. We will consider versions of substance dualism that recognize a much more intimate link between consciousness and neurophysiology in chapter 6.

53. Note that this objection is aimed at only substance dualism, and not property dualism. According to the latter position, nonphysical, psychological states are indeed properties or states of the body.

William Hasker has argued that this response by the dualist falls far short: "But why should *consciousness itself* be interrupted by a blow to the head, or a dose of medication? And why should a personality be drastically altered—sometimes temporarily, sometimes permanently—by injury to the brain or a chemical imbalance in the brain?"[54] Hasker's point is that it seems that drugs, injuries, chemical balances, and such affect not merely the information given to a mind, but also the character of the mind itself and even the presence of mind itself in the cases of loss of consciousness. It doesn't appear that when one has had too much to drink he is fully capable of having a normal conscious experience behind the scenes, as it were. Moreover, it is quite far-fetched, once again, for the dualist to claim that despite all appearances otherwise, Phineas Gage's mind continued to reside "in" his body unharmed, despite the radical changes in his personality. Thus, Hasker concludes that "an adequate theory of mind must allow for the dependence of mental activity upon brain function in a way that is stronger than Cartesian dualism can readily accommodate."[55]

Other philosophers worry that dualism is unbelievable given certain plain facts revealed by our common experience of living as human beings. I don't experience myself primarily as a being who happens to occupy a body for the time being, but as an embodied person, literally an organism of a certain type involved in physical relations with other physical beings. My relationship to my body is not the same as my relationship to the computer keyboard. The keyboard, as something entirely external to my existence, like any other tool, is merely an instrument I act on when pursuing certain ends. At least given common experience, there

54. William Hasker, "Persons as Emergent Substances," in *Body, Soul, and Survival: Essays on the Metaphysics of Human Persons,* ed. Kevin Corcoran (Ithaca, N.Y.: Cornell University Press, 2001), 112. Author's emphasis.

55. Ibid. Hasker is himself a substance dualist, but his position is that the mind or soul emerges from the body in a way that avoids the problems he raises. We will discuss emergentism in chapter 6.

is some sense in which I am a human body. If I have a good day in the gym and deadlift six hundred pounds, it's not my body as directed by my mind, but *me* who lifts the weight. It isn't merely my body that is hungry, tired, energetic, and so on, but *I* am in such states. Moreover, it isn't my body alone that acts by breathing, eating, sleeping, or digesting, but I do. Human beings are a certain kind of animal, and like other animals our bodies are not merely external instruments, but rather they are essential to what we are. Robert George makes this point well when he claims that

> the dualistic view of the human person makes nonsense of the experience all of us have in our activities of being dynamically unified actors—of being, that is, embodied persons and not persons who merely "inhabit" our bodies and direct them as extrinsic instruments under our control, like automobiles. We don't sit in the physical body and direct it as an instrument, the way we sit in a car and make it go left or right.[56]

In other words, experience of ourselves in our ordinary acts of agency runs contrary to the claim that we are contingently embodied minds that might just as easily have existed without bodies. In fairness to the substance dualist, even Descartes warns against thinking of the soul as being related to the body just as a captain is related to a ship, but it is difficult to see just how the substance dualist can resist such an external-instrumental account of mind-body relations.[57]

Along these lines, critics of substance dualism are worried about not only the counterintuitive phenomenology entailed by dualism, but also its apparent moral consequences. The concern here is that dualism, since it takes the body as an external instrument of the self, seems to lead us to conclude that the human body is a *subpersonal* entity, to borrow George's phrase, whose moral standing is less than that of persons. If I am not my body,

56. Robert George, *The Clash of Orthodoxies: Law, Religion, and Morality in Crisis* (Wilmington, Del.: ISI Books, 2001), 9.

57. See Descartes, *Meditations on First Philosophy,* in *The Philosophical Writings of Descartes,* 2:56. John Cottingham argues for a more subtle picture of Descartes's philosophy of mind in "Cartesian Trialism," *Mind* 94, no. 374 (1985): 218–30.

as substance dualism seems to imply, the human body has a low-grade moral standing no different in principle from the value of other external instruments or tools. That is, on a substance dualist understanding, we would expect that human bodies lack intrinsic value, and are valuable only to the degree that they are useful to or desired by the minds that contingently occupy them. Since life is a biological process of a human body, it would then seem that human life, like the body, has only instrumental value. That is, there is nothing intrinsically valuable about the life of the body, since the body is merely a means to the ends of the mind or soul. George argues that "it is the status of the body as subpersonal that accounts for the willingness … to authorize the killing of human beings before they become persons … and after they cease being 'persons' because only persons 'have dignity and rights' and merely 'living bodies' do not."[58] One should be concerned about a theory of human nature that cheapens the value of human life in general and the lives of those most vulnerable among us in particular. Thus, given both the experiential and moral considerations we have discussed, we do well to regard ourselves as bodily beings.[59]

Finally, note the obvious fact that human beings are not the only animals that have psychological states. There can be little doubt that other animals have sensations; for example, Fido the dog clearly feels pain when a child pulls at his tail, and he experiences the taste of a piece of sausage he has begged for. Likewise, there is some reason to believe that higher animals, for example, apes and dolphins, have thoughts in some very broad sense, even if primitive.[60] Of course it is possible that animals lack the inward

58. George, *The Clash of Orthodoxies,* 35.

59. See Robert George and Patrick Lee, *Body-Self Dualism in Contemporary Ethics and Politics* (New York: Cambridge University Press, 2009) for a sustained critique of dualism along metaphysical, moral, and phenomenological lines. Charles Taliaferro defends substance dualism against moral critiques like George's in "The Virtues of Embodiment," *Philosophy* 76, no. 295 (January 2001): 111–25.

60. See chs. 3–5 of Alasdair MacIntyre, *Dependent Rational Animals: Why Human Beings Need the Virtues* (Chicago: Open Court , 1999).

experience of sensation while presenting the appearance, but the same could be said for other people. It is possible that I'm the only being with sensations and thoughts, despite appearances otherwise, but few people would take such a thesis seriously. Thus, it is very hard to take the claim that animals lack sensations seriously. If we accept the arguments for substance dualism, are we then to conclude that the psychological states of nonhuman animals are likewise possessed by immaterial substances? That's certainly logically possible, but we are now being asked to believe in quite a few immaterial substances.[61] It seems then that we have some reason to doubt the inference the dualist makes from sensations and thoughts to nonidentity or independence, unless we are willing to do the same in the case of all conscious animals.

None of these objections is ultimately decisive against dualism, and the same can be said regarding the various objections regarding mind-brain interaction. In each case there is some avenue for the dualist to reply. Dualism, including substance dualism, is not defeated; it is very much a live option for our understanding of the mind. The problem for the substance dualist, however, is that to maintain her position, she needs to tell increasingly complicated, and in some cases far-fetched, stories, for example, different occurrences in the brain in nonexperimental contexts, subtle reinterpretations of fundamental laws of nature, the presence of minds even in humans that can no longer give outward evidence of psychological states, a counterintuitive phenomenology, nonphysical animal minds, and so on. Dualism seems to add up to a very complicated and at times counterintuitive theory. *Counterintuitive* is not the same as *incoherent.* The former is just hard to believe, whereas the latter is impossible to believe. The materialist's objections do not succeed in showing dualism to be incoherent, but to be counterintuitive and theoretically complicated.

61. Richard Swinburne, a substance dualist, believes that animals do have Cartesian-style minds. See Swinburne, *The Evolution of the Soul,* rev. ed. (New York: Oxford University Press, 1997), especially ch. 10.

Less complication is not necessarily better, if a more complicated theory is necessary to explain the relevant facts, and nothing bars reality from turning out to be an odd place.

We could then settle for some version of dualism, if in the end it is the only viable option for explaining sensations and thoughts. If other theories are available that lack these complicating factors while making sense of sensations and thoughts, then these theories would be preferable to dualism. Since materialism takes human beings to be physical substances the same in kind as other physical substances that possess only physical properties, problems regarding mind-brain interaction, the dependence of psychological states on neurophysiological states, our experience of ourselves as embodied beings, and the psychological status of nonhuman animals do not arise for the materialist. Remember that the materialist is out to execute a two-phase strategy: (a) to show dualism to be false or at least highly improbable; and (b) to show that premise (2) of the difference argument (the mind has essential properties that the brain essentially lacks) is false. In the foregoing discussion we have not (by a long shot) shown dualism to be absurd, but we may have complicated it enough to support a weakened first phase of the materialist's two-phase strategy; though dualism is not unbelievable, it is a position of last resort. If we could give good reason to believe that our possession of psychological properties does not imply that we have nonphysical properties, then the second phase of the materialist strategy would be completed. We would then likewise have just enough evidence to reject dualism and accept materialism.[62] In the following chapter we will discuss a series of influential attempts by materialists to complete the second phase of this strategy.

62. Remember that for the time being we are assuming that our only option is either materialism or some version of dualism. This is an assumption we will question in chapters 6–8 below.

CHAPTER 4

MATERIALISM

In the previous chapter we discussed a two-phased materialist strategy for defeating dualism. In the first phase, the materialist argues that dualism is false (or at least beyond reasonable belief). In the second phase, the materialist either argues that there are no such things as psychological states or constructs an account for our psychological states that does not take sensations and thoughts as nonphysical entities. We found that the first phase is not complete in any strong sense; that is, we don't have good reason to conclude that dualism is beyond the pale. We did find, however, that the dualist (especially the substance dualist) does need to complicate matters greatly in order to reply to these arguments, so a weakened first phase of the materialist project can be defended; that is, we should accept dualism only in the absence of any better account of sensations and thoughts. The materialist can show that there is just such an account by successfully executing the second phase of his project, which is to defend an entirely materialist understanding of our psychological states (or to show that it is plausible to suppose that there are no such things

as psychological states). We will spend this chapter discussing the prospects of this second phase of the materialist's strategy.

Eliminative Materialism

Among the various forms of materialism, the easiest version to understand is *eliminative materialism.* To say that we eliminate *x* in favor of *y* is to claim that *x* and *y* are not identical and we now have replaced *x* with *y,* in the sense that we no longer believe there is any such thing as *x* and instead believe in the existence of *y.* For example, in early modern theories of combustion, phlogiston was posited to account for the loss of mass in burned materials. The idea was that some material, phlogiston, must be emitted by the burning substance. Since phlogiston seemed to explain combustion (along with other phenomena including the rusting of metals), scientists believed there was such a material, even though it was not thought to be directly observable. The phlogiston theory was later found to have internal problems, and it was disconfirmed experimentally. A replacement theory, positing oxygen as something consumed during combustion, made better predictions and was free of internal difficulties, while accounting for the same phenomena. Thus, there is no longer reason to believe in phlogiston, but we do accept the existence of oxygen. In short, phlogiston has been *eliminated* in favor of oxygen in our theory of combustion.

Note that in a case of elimination, we don't claim to find out that the object eliminated is identical to or turns out to be the same thing as the replacement theoretical entity. Scientists do not accept the following proposition:

(1) Phlogiston is oxygen.

We have discovered not that phlogiston *is* oxygen, but that *there is no such thing as phlogiston*! Rather we now know that something entirely different accounts for combustion, namely oxygen. Theoretical eliminations do not result in true identity propositions, but the replacement of a false set of propositions (those postulat-

ing entities of the now obsolete theory) with what we now take to be a true set of propositions (those postulating the entities of the new theory). Since there is no such thing as phlogiston, most propositions that refer to it are strictly false. For example,

(2) The burning grass emitted phlogiston

is straightforwardly false, because there is no such thing as phlogiston. If you persist in a phlogiston theory of combustion, you maintain a set of false beliefs.

Some materialists propose an eliminativist account of psychological entities that literally denies the existence of sensations and thoughts just as we now deny the existence of phlogiston. Paul Churchland, who is the foremost defender of eliminative materialism, defines his position as follows: "Eliminative materialism is the thesis that our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and ontology of that theory will eventually be displaced … by a completed neuroscience."[1] According to Churchland and other eliminativists, sensations and thoughts are the entities of a primitive theory (so-called folk psychology) we all employ to explain each other's outward behavior. We see and hear certain activities on the part of other organisms, and we then posit unseen, inner episodes, that is, sensations and thoughts, to explain these phenomena. Eventually we become so skilled at applying this theory that we use it to explain our own behavior.[2] The eliminativists argue that we now have a better theory provided by contemporary neuroscience that can

1. Paul Churchland, "Eliminative Materialism and Propositional Attitudes," *Journal of Philosophy* 78, no. 2 (1981): 67–90. For an early defenses of eliminative materialism, see Richard Rorty, "Mind-Body Identity, Privacy, and Categories," in *Materialism and the Mind Body Problem,* ed. David M. Rosentthal (Indianapolis, Ind.: Hackett Publishing, 2000), 174–99, and Paul Feyerabend, "Mental Events and the Brain" in the same volume, 172–73.

2. Wilfrid Sellars offers a detailed and influential account of folk psychology as a primitive theory in "Empiricism and the Philosophy of Mind." Note that Sellars himself does not believe that folk psychology is as yet a refuted theory, and he definitely does not accept eliminative materialism. His most sustained attempt to preserve folk psychology is "Philosophy and the Scientific Image of Man," in *Science, Perception, and Reality* (Atascadero, Calif.: Ridgeview Publishing, 1991), 1–40.

explain our outward behavior without positing psychological episodes. Since we have a better theory, we can see that *"our commonsense psychological framework is a false and radically misleading conception of the causes of human behavior and the nature of cognitive activity,"* and "we must expect that … the older framework will simply be eliminated … by a matured neuroscience."[3]

Thus, for the eliminative materialist,

(3) Jack is in pain

and

(4) Brendan thinks he is handsome

are strictly *false.* There are no such events or entities as sensations (e.g., pains) and thoughts (e.g., thinking oneself to be handsome). Instead, we now know that

(5) John's neural-network *X* is activated

and

(6) Brendan's neural-network *Y* is activated

are the true descriptions of what occurs when we typically utter the falsehoods (3) and (4). The case for eliminativism is basically the fruitfulness of contemporary neuroscience, which can seemingly operate without reference to psychological states, and the supposed explanatory sterility of folk psychology. Neuroscience is producing a progressively sophisticated and satisfying picture of how human cognition works, whereas, so argues the eliminative materialist, folk psychology terminates in obscurities, far-fetched stories, and seemingly unintelligible notions associated with dualism. We do well then to reject folk psychology as a failed theory (though it may still be useful for daily life in the way that the false statement "The sun rises in the east" is practically useful), and instead adopt the theoretical vocabulary of neuroscience as our idiom for speaking about the mind.[4]

3. Paul Churchland, *Matter and Consciousness,* 2nd ed. (Cambridge, Mass.: MIT Press, 1988), 43. Author's emphasis.

4. Paul and Patricia Churchland present detailed arguments to this effect in "Folk

Life would be easy for the materialist if eliminativism were true. There would be no need to explain psychological phenomena in any way, because all talk of psychological phenomena would be eliminated from our theoretical understanding in favor of neurological descriptions. However, things do not turn out so neatly for the materialist. First of all, it is far from clear that folk psychology is really a theory, and not a direct observation. Do we posit sensations and thoughts to make sense of behavior in the way we posit subatomic particles to make sense of observable phenomena? That is, some philosophers have argued that our thoughts and sensations are not theoretical entities that we posit for explanatory purposes, but the very data that need to be explained by a proper psychological theory.[5] The apparently incorrigible, direct access we have to our own mental lives counts heartily against attempts to argue that we posit inner episodes in order to explain outward behavior in ourselves and others.[6] Moreover, even if one grants that folk psychology is indeed a theory, then it is likely among the oldest and most successful theories in human history. It seems at least prima facie improbable that such a wildly successful theory would turn out to be utterly off the mark.[7]

One might attempt to reply on behalf of the eliminativist by

Psychology" in *On the Contrary: Critical Essays 1987–1997* (Cambridge, Mass.: MIT Press, 1998), 3–17.

5. Flanagan, *Consciousness Reconsidered,* 26–27. William Hasker discusses Flanagan's and similar criticisms of eliminativism in *The Emergent Self,* 4–5. John Searle gives a nontheoretical account of folk psychology contrary to the eliminativist view in *The Rediscovery of the Mind* (Cambridge, Mass.: MIT Press, 1992), 58–64.

6. Sellars attempts to account for our apparent direct access to our sensations and thoughts by claiming that we have become so apt at applying folk psychology to ourselves that we no longer notice that we reach inner episodes by an inference and not direct observation. See "Empiricism and the Philosophy of Mind," 186–89, 194–96. Something like the "historical fiction" that Sellars uses to make this point might very well be possible, but one is still left to wonder how likely it is that I don't have direct access to my own thoughts. Many philosophers (e.g., see Searle in the previous note) consider these sorts of positions simply to strain credulity.

7. See Terence Horgan and James Woodard, "Folk Psychology Is Here to Stay," *Philosophical Review* 94, no. 2 (1985): 197–220, along with Hasker's discussion in *The Emergent Self,* 4–5.

arguing that our supposed incorrigible access to our psychological states is a systematic illusion we fall into because of the long history and practical success of folk psychology. In the past human beings have been prey to systematic illusions as an effect of their mistaken, though maybe practically helpful, theories, which have at times been dispelled by subsequent scientific discovery. One example frequently cited by eliminativists is that people used to think they really experienced being possessed by demons or suffering from spells cast by witches. Even though it really seemed to premodern people that they were witnessing (or even enduring) a possession, today we now know that they, for the most part at least, suffered from mental illnesses such as schizophrenia, not actual demon possession.[8] Our contemporary sciences now show us that many people who thought they were experiencing demonic possession were really probably experiencing some neurophysiologically based mental illness. They were simply wrong about what they were experiencing, and we have since rightly *reconceived* the sort of events that used to be thought of as cases of demonic possession in terms of mental illness. Likewise, says the eliminativist, we would do well to reconceive our illusions of sensations and beliefs in a properly neurological way. As Richard Rorty, an early defender of eliminative materialism, puts it, "we have just been reporting neurons when we thought we were reporting raw feels [i.e., sensations] … but once we were clued in we could resdescribe what we had been reporting easily enough."[9] A worry for the eliminativist is that it is unclear how our discovery of our prior illusions and our newly enlightened reconceptualization or redescription is even possible, because "Conceptions and illusions are intentional entities. So if [eliminative materialism] is true, we can neither reconceive phenomena nor discover that we had been harboring the illu-

8. Rorty discusses this sort of example at length in "Mind-Body Identity, Privacy, and Categories," 178–81.

9. Rorty, *Philosophy and the Mirror of Nature,* 83.

sion that there were intentional states."[10] That is, it is unclear as to how any of this talk of having "illusions" and "false beliefs" about our supposed psychological states along with the notion of a "reconceptualization" using the "concepts" of neurology makes any sense, if we suppose that there are no such things as psychological states, because this idiom implies that we have psychological states.

An even worse problem looms in this vicinity. The eliminativist *asserts* her theory as *true.* Assertion and truth are inherently intentional or psychological notions, with which the eliminativist would have us dispense; the very notion of affirming or denying a theory as true is a psychological process that stands to be eliminated. To *assert* as *true* a theory that denies the reality of assertion and truth is at least a practical impossibility, if not outright contradictory. Along these lines some critics then argue that eliminative materialism is a *self-refuting* theory; that is, to affirm it is, at least implicitly, to deny it.[11] In other words, the problem for eliminative materialism is that any rejection of folk psychology "will ask us to accept the conclusion that [folk psychology] is false, but the very notion of acceptance is one that still presupposes [folk psychology]."[12] One frequent reply to the charge of self-refutation is to recommend that we replace our traditional conceptions of assertion and truth in favor of a *pragmatism* that would have us think of truth in terms of usefulness or adaptability instead of getting the world right or correspondence to reality.[13] One worries, however, whether pragmatism is

10. Angus Menuge, *Agents under Fire: Materialism and the Rationality of Science* (New York: Rowman and Littlefield, 2004), 48.

11. For a detailed version of this argument, see Lynn Rudder Baker, *Saving Belief: A Critique of Physicalism* (Princeton, N.J.: Princeton University Press, 1987).

12. Menuge, *Agents under Fire,* 46. William Hasker defends the self-defeating objection to eliminative materialism in *The Emergent Self,* 5–20.

13. See Paul Churchland, "The Ontological Status of Observables," in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science,* 139–51 (Cambridge, Mass.: MIT Press, 1989). Hasker critically discusses this proposal in *The Emergent Self,* 14–17.

itself wrought with paradox; that is, it clearly cannot be asserted as *true* in the traditional sense, and given the practical success of folk psychology it does not seem likely that eliminativism is to be preferred on pragmatic grounds.[14]

Self-referential incoherence is about as bad a consequence as a theory can entail, so we will assume that any materialist theory of mind that slips over the cliff into eliminativism fails. We might then think of the materialist's task as trying to steer between eliminativism and dualism; that is, the materialist must find a way to affirm the reality of sensations and thoughts (thus avoiding eliminativism) that renders such phenomena straightforwardly physical (thus avoiding dualism).

Behaviorism

We should not confuse logical behaviorism with the behaviorist school of psychology associated with B. F. Skinner. There are certainly similarities between the two views, but psychological behaviorism does not purport to be a metaphysical theory of mind at all. Though it has fallen out of favor even among materialist philosophers of mind, logical behaviorism enjoyed a great deal of influence in the first half of the twentieth century, and a close consideration of this theory will give a good sense of how other forms of materialism work.[15] Carl Hemple, one of the early proponents of this doctrine, defines logical behaviorism as the view that "the meaning of a psychological statement consists

14. Richard Rorty proposes a pragmatic approach to truth in *Philosophy and the Mirror of Nature* and more explicitly in later works (see *Contingency, Irony, and Solidarity*). He also grants that this turn forces us to jettison anything like the notion of truth corresponding to reality.

15. The most well-known presentation of logical behaviorism is Gilbert Ryle's *The Concept of Mind* (Chicago: University of Chicago Press, 1949). For another influential, though more technical, version of behaviorism see also Rudolf Carnap, "Psychology in Physical Language," *Erkenntnis* 3 (1932–33): 107–42, reprinted in *Logical Positivism,* ed. A. J. Ayer, (New York: Free Press, 1966), 165–98, though in this article Carnap ultimately conjectures what he takes to be a better though nonbehaviorist version of materialism. Sellars offers a critical analysis of a number of forms of logical behaviorism in "Empiricism and the Philosophy of Mind."

solely in the function of abbreviating the description of certain modes of physical response characteristic of the bodies of men and animals."[16] Consider the following proposition:

(7) Will is in pain.

This proposition attributes to Will a certain mental property, *being in pain*. Normally, we would think of (7) as a description of an internal, private state had by Will. If Will were to say, "I'm in pain," we would take this as a *report* on his internal, psychological state, similar to somebody's reporting a state of affairs from behind a closed door to people outside the room. As we have discussed at length, the materialist has a choice of either eliminating reference to Will's sensations and thereby taking (7) to be false, or accounting for pain as something physical. The behaviorist proposes to define pain in terms of behavior: Will's *being in pain* just is his overt physical behavior. Thus, claims the behaviorist, (7) can be translated to

(8) Will is moaning, wiggling, suffering a rise in blood pressure, and so on, following a bodily injury.

In short, the behaviorist proposes a translation of all sentences mentioning sensations to equivalent sentences mentioning only behavior. Note that (8) makes no mention of psychological states at all, just observable bodily behavior. Since bodily behavior is straightforwardly physical, if this translation is successful, we would seem to have an impeccable materialist account of sensations. For the behaviorist, if Will were to say, "I'm in pain," he would not make a report on his internal state; his utterance would just be a very complicated form of pain-behavior, a sophisticated type of moaning or wincing. The behaviorist will also try to motivate this theory by pointing out that it is only following an observation of somebody's outward behavior that we claim him or her to be in a psychological state. Thus, it seems that our language of

16. Carl Hemple, "The Logical Analysis of Psychology," in *The Philosophy of Mind: A Guide and Anthology,* ed. John Heil, (New York: Oxford University Press, 2004): 91.

psychological states really refers to behavior, not some unseen inner events. According to the behaviorist, we could switch (8) for (7) without changing the meaning of the proposition. Pain *just is* the wiggling, moaning, cursing, changes in blood pressure, and so on that mark the proper use of this concept.

Note well that *the behaviorist, unlike the eliminativist, does not claim that* (7) *is false* (*she doesn't deny that Will is in pain*), *but she claims that what we really mean by* (7) *is* (8)*; to be in pain just is to engage in certain modes of behavior.* As Hemple puts it, "Logical behaviorism claims neither that minds, feelings, inferiority complexes, voluntary actions, etc. do not exist, nor that their existence is doubtful," though it does assert that "the propositions of psychology are consequently physicalist propositions."[17] That is, the logical behaviorist grants that we have psychological states, but on her reckoning they are physical phenomena because they can be identified with overt, in principle empirically observable, behaviors. Thus, logical behaviorism can be taken to affirm both materialism and the reality of psychological states, because the latter just are physical entities or events, that is, certain modes of bodily behavior.

The behaviorist might attempt to analyze more complex cases such as

(9) Martha sees a truck coming

as

(10) Martha is standing on the sidewalk in the presence of an approaching truck under normal visual conditions, and so on.

Of course (10) is a greatly simplified version of what would be a highly complex analysis, but it should give us the flavor of how behaviorism works. Notice that (10) makes no use of typical psychological terms ("sees" drops out), and it describes the conditions under which we would assert (9). Thus, concludes the behaviorist,

17. Ibid., 92 and 91.

(10) is a proper analysis of (9), and we can therefore reduce Martha's *seeing* to her behavior.

We can envision a similar behaviorist analysis of propositions attributing beliefs. Consider a proposition about a belief:

(11) Will believes that it will snow.

In cases such as (11) the behaviorist can often appeal directly to the agent's behavior:

(12) Will put his boots on before leaving the house.

The appropriate behavior for believing it will snow is putting one's boots on, or some other such activity. Other beliefs are a bit more difficult, because we hold them even in cases in which we are not behaving in any way characteristic of such a belief. For example,

(13) Patrick believes that 9 x 9 = 81.

Assuming that Patrick is arithmetically competent, it is true in some sense that he believes that 9 x 9 = 81, even when he is not currently behaving in an overt manner that betrays such a belief, for example, at those times at which he is not uttering "9 x 9 = 81." In such cases, the behaviorist introduces the notion of a *disposition.* When we say that salt has a disposition to dissolve in water, we mean that *if salt is introduced into water under standard conditions, it will dissolve.* Our dispositions are what we would do, if certain circumstances obtain. The behaviorist may then analyze (13) as

(14) Patrick is disposed to reply "Yes," when he is asked "Does 9 x 9 = 81?"

To have a belief, according to the behaviorist, is to have a certain disposition to make either a verbal or nonverbal response in a certain situation. Beliefs are like loaded springs; the person holding the belief is poised to act or speak in a certain way when appropriate conditions arise. Note once again that (14) manages to analyze (13) without any remaining reference to psychological states.

It is also important to note that in none of these analyses has the behaviorist claimed to eliminate or "get rid of" thoughts or sensations. Rather the behaviorist claims to have shown that Will's pain *just is* his wiggling and moaning, Martha's seeing *just is* her remaining in place in the face of the oncoming truck, and Patrick's belief *just is* his long-term disposition to answer a certain question affirmatively. In other words, the behaviorist claims to have given us a theory of the meaning of psychological terms; phrases such as "is in pain" and "believes that the sky is blue" really mean the same thing as descriptions of overt behavior and behavioral dispositions. As long as the behaviorist can come up with an exhaustive set of analyses translating our mental language into physical language referring only to overt behavior or behavioral dispositions, she believes she has reduced psychological states to physical states without giving up common-sense truths about beliefs, thoughts, feelings, and the like. In short, if behaviorism works, then one may grant the reality of psychological states without postulating anything mental or nonphysical, by reducing thoughts and sensations to behaviors and dispositions for behaviors.

Behaviorism is the subject of withering criticism, from dualists and materialists alike.[18] Of particular note, the concept of a disposition is notoriously difficult to analyze in a way that makes its application anything more than trivial. If we say that the glass broke in a certain circumstance because it was brittle, and "brittle" just is the disposition to break under these particular conditions, our explanation is merely trivial. Likewise if we say that Patrick answered "Yes" because he believes that "9 x 9 = 81," and we analyze "Patrick believes that '9 x 8 = 81'" as "Patrick is disposed to answer 'Yes' when asked 'Does 9 x 9 = 81'," then we have likewise made a trivial claim.[19]

18. Searle offers a concise and helpful discussion of all the standard objections to behaviorism in *The Rediscovery of the Mind,* 34–35. His way of framing these issues has certainly influenced what follows in my discussion of the criticisms of behaviorism.

19. See Peter Geach, *Mental Acts* (London: Routledge and Kegan Paul, 1957), 4–7.

There is also a problem of circularity for the behaviorist. Consider once again (9) and (10) from above:

(9) Martha sees a truck coming.

(10) Martha is standing on the sidewalk in the presence of an approaching truck under normal visual conditions.

The analysis of (9) by (10) is plausible only if we accept the truth of a further proposition:

(15) Martha does not *want* to be hit by the truck.

The problem here is that we have now analyzed one psychological term, "sees," using another, "wants." We are just trading one psychological term referring to a sensation for another referring to an intention. The logical behaviorist might try to slip out of this by analyzing (15), but the most likely behavioral analysis is

(16) Martha is standing on the sidewalk in the presence of an approaching truck under normal visual conditions.

That is, our analysis of thinking that Martha does not want to be hit by a truck is the same that we used for (10), but (15) was supposed to be part of an analysis of the meaning of (10), so we are now going around in a circle. As Searle puts it, circularity is an unavoidable problem for many behaviorist analyses, because "to give an analysis of belief in terms of behavior, it seems that one has to make reference to desire; to give an analysis of desire, it seems that one has to make reference to belief."[20] It seems then that in many cases the behaviorist cannot give a noncircular analysis that makes no reference to a psychological state distinct from overt behavior or behavioral dispositions, so we don't have solid reasons to accept the behaviorist's claim that psychological states just are overt behaviors and behavioral dispositions.

Leaving aside what Searle calls technical objections to behaviorism, we can also readily see that behaviorism violates some of our most obvious intuitions about mind. In the last chapter, we concluded that a theory of mind that denies the possibility

20. Searle, *The Rediscovery of the Mind,* 34.

of psychological states causing our behavior is unacceptable. Behaviorism falls into such an epiphenomenalism.[21] Consider the following proposition:

(17) Patrick is crying because he is in pain.

Here we are given an explanation of Patrick's behavior (crying) in terms of his psychological state (pain). This kind of explanation of overt behavior in terms of a psychological state is quite familiar to us and is certainly plausible. Note that in (17) we don't claim that Patrick's pain *is* his crying, but that Patrick's pain *causes* his crying. The behaviorist would have to analyze (17) as follows, because she believes that *Patrick's pain just is his crying:*

(18) Patrick is crying because he is crying.

The behaviorist's (18) is trivial, and it would likewise make all claims of mind-body causation trivial. There is no room for mental causation in a behaviorist theory of mind. Hilary Putnam makes this point well: "both the dualist and materialist would want to argue that although the meaning of 'pain' may be explained by reference to overt behavior, what we mean by 'pain' is not the presence of a cluster of responses, but rather the presence of an event or condition that normally causes those responses."[22] In other words, sensations and thoughts are not behavior (or behavioral dispositions), but causes of behavior. To conflate them with behaviors is to undermine our cornerstone understanding of psycho-physical causation. Behaviorism then slides into an unacceptable form of epiphenomenalism.

Finally, there is just something flatly unbelievable about behaviorism. If the behaviorist is correct, then our psychological states just are our overt behaviors. That implies that other people

21. Actually, behaviorism might leave the possibility that our psychological states do have effects on other people (e.g., my pain behavior might elicit a response from an onlooker), but as we will see below it cannot account for how *my* psychological states can be the cause of or explanation for *my* behavior, which is the more worrisome consequence of epiphenomenalism. Thanks to a blind reviewer for making this distinction.

22. Hillary Putnam, "Brains and Behavior," in *Philosophy of Mind: Classical and Contemporary Readings,* ed. David Chalmers (New York: Oxford University Press, 2002), 47.

are just as well, if not better, acquainted with our psychological states than we are ourselves. Since you can better see the expression on my face, hear the intonation of my voice, and so on, you might be in a better position to know how I am feeling than I am! For the behaviorist, we have no privileged access to our own psychological states. There is no first-person perspective; psychological states are just as third-person available as any other physical state. What about very good liars, gifted actors, and soldiers trained to ignore pain flawlessly? How about our own internal sense of our privileged access to our own thoughts and feelings? Of course there are times when others know our internal states better than we do, but to say that we have no better access to our mental states than any third-person observer is to strain credulity to the breaking point.

Identity Theory

The behaviorist claimed, quite ambitiously, that psychological terms are translatable into terms referring to overt behavior or behavioral dispositions. That is, psychological terms were supposed to be logically equivalent to corresponding behavioral terms. If that is correct, when we use psychological terms, we are actually talking about overt behavior. As we have seen above, the behaviorist's semantic thesis is clearly off the mark. Many materialists, in light of the failures of behaviorism, admit that psychological terms are not logically equivalent to terms referring to physical states of affairs, especially behaviors. They argue, however, that conceptual or logical differences do not make for real distinction between entities. The product of this argument is commonly called the *identity theory.*[23] Almost all forms of materialism taken seriously for the last several decades are some variation of the identity theory as originally developed and defended

23. What we are calling the identity theory has in fact been given many different names in the philosophical literature. For a useful taxonomy, see Thomas W. Polger, "Identity Theories," *Philosophy Compass* 4, no. 5 (2009): 822–34.

by U. T. Place, H. Feigl and J. J. C. Smart.[24] We will follow Smart's version of the theory initially.

The identity theory makes heavy use of what I will call *empirical identity.* Some identity statements can be known without doing any empirical investigation, what philosophers call "a priori knowledge." The term *a priori* is Latin for "prior to," and in this case, it signifies knowledge that can be had prior to any experience. For example the following proposition is an a priori identity:

(19) Jim Madden is Jim Madden.

You don't need to investigate anything empirically to know that (19) is true. Moreover, (19) is certainly true in all possible worlds, that is, it would be a logical contradiction to say that "Jim Madden is not Jim Madden," so it cannot be false in any situation. Likewise, (20) does not raise any empirical questions:

(20) The bachelor I spoke to yesterday is the unmarried man I spoke to yesterday.

Even though "bachelor" and "unmarried man" are different phrases, they express the same concept, that is "bachelor" and "unmarried man" amount to the very same thing, so we need do no more than inspect the meanings (concepts) of the terms involved to see that this is a true identity statement.

24. See U. T. Place, "Is Consciousness a Brain Process?" *British Journal of Psychology,* 47 (1956): 44–50; Herbert Feigl, "The 'Mental' and the 'Physical,'" in *Concepts, Theories, and the Mind-Body Problem,* ed. H Feigl, G. Maxwell, and M. Scriven, 370–497, Minnesota Studies in the Philosophy of Science 2 (Minneapolis, Minn.: University of Minnesota Press, 1958); and J. J. C. Smart, "Sensations and Brain Processes," *Philosophical Review* 68 (1959): 141–56. I will refer to the version of Place's paper that appears in Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings,* 55–60; and the version of Smart's paper that appears in *Materialism and the Mind-Body Problem,* 2nd Edition, ed. David M. Rosenthal, 53–66 (Indianapolis, Ind.: Hackett Publishing, 2000). David Armstrong provides a very helpful commentary on both Place's and Smart's papers in *The Mind-Body Problem,* 67–81. Polger, "Identity Theory," likewise provides a very useful starting point. Place and Smart are certainly innovative in their use of the distinction between conceptual and real distinction in defense of materialism, but Carnap clearly has this sort of reduction in mind in his earlier "Psychology in Physical Language," though he can only conjecture (he settles for a "stop-gap" behaviorism) as he writes before the introduction of certain conceptual tools and the neuroscientific progress necessary to develop it.

Consider, however, "the Morning Star" and "the Evening Star." Notice that these phrases express two distinct concepts—"the Morning Star" is conceptually distinct from "the Evening Star" in the sense that one doesn't need to know a single thing about what it means to be the Evening Star in order to be aware of the Morning Star. In fact, one could be well acquainted with the Evening Star while even doubting the existence of the Morning Star, and vice versa. It turns out, however, that the Morning Star and the Evening Star refer to one and the same thing in the actual world, that is, the planet Venus. Thus, the following is a true identity statement:

(21) The Morning Star is the Evening Star.

The point here is that conceptual distinction does not imply any actual distinction; one and the same thing can be accurately described by two very different concepts. In such cases mere a priori considerations would not reveal the identity. Rather, only empirical investigation reveals that what is conceptually distinct is numerically identical. Indeed, until empirical discoveries revealed otherwise, ancient astronomers believed that the Morning Star and the Evening Star were distinct objects.

Identities like (21) are often called *contingent identities,* because there are possible worlds in which there is an entity that has the properties sufficient to be the Morning Star and lacks the properties necessary to be the Evening Star (and vice versa), though in the actual world it has worked out such that the same entity has both sets of properties.[25] It would *seem* then that (21) is a contingent truth, because there are worlds in which it is false. We need to be careful here, however, because we need to make a fairly picky linguistic distinction, which I mentioned briefly in chapter 2. If we take "the Morning Star" and "the Evening Star" to be ordinary predicates that indicate properties (or sets of properties), that is, *being the Morning Star* and *being the Evening*

25. The discussion in the next two paragraphs is rather technical. The more casual reader might skip it without losing the essentials of this section.

Star, then (21) is contingent. Here is a contingent interpretation of (21):

(21*) The entity that has the properties sufficient to be the Morning Star also has the properties sufficient to be the Evening Star.

Notice that (21*) simply claims that one entity has two different sets of properties that are only contingently associated. On the other hand, we could interpret "the Morning Star" and "the Evening Star" as what philosophers of language, following Saul Kripke, call *rigid designators.* In this case "the Morning Star" and "the Evening Star" are not predicates but *names* or *referring expressions* that merely point to an object without conveying information about it, and here we assume that the referent of these terms remains constant across all possible worlds.[26] Thus, the following proposition is a necessary truth:

(21**) The entity referred to by "the Morning Star" is identical to the entity referred to by "the Evening Star."

Since our reference for the two rigid designators remains constant in all of the various possible worlds, (21**) is necessarily true. The referent of both terms is Venus, and "Venus is Venus" is true in all possible worlds. Thus, (21**) is necessarily true, and for this reason we will call propositions of this sort *empirical identities,* rather than contingent identities.

Empirical identities abound in the results of natural science. For example,

(22) Water is H_2O.

and

(23) Lightning is an electric discharge.

26. Kripke's seminal discussion of this issue appears in Lecture I of *Naming and Necessity.* See also Hillary Putnam's "Is Water Necessarily H2O?" in *Realism with a Human Face,* ed. James Conat (Cambridge, Mass.: Harvard University Press, 1990), 54–79. Scott Soames provides an excellent introduction to these issues in *Philosophical Analysis in the Twentieth-Century,* vol. 2 (Princeton, N.J.: Princeton University Press, 2003), 333–456. See Gary Gutting, *What Philosophers Know: Case Studies in Recent Analytic Philosophy* (New York: Cambridge University Press, 2009), 31–50, for an accessible discussion of these issues that is not entirely sympathetic to Kripke's position.

We know that the clear, tasteless liquid that constitutes two-thirds of our bodies is identical to collections of H_2O. However, you could recognize water phenomenally without identifying it with H_2O. People were well acquainted with water for millennia without an inkling that it is H_2O, just as my five-year-old son, Brendan, can correctly identify a quantity of water, but he has no clue that it is identical to a quantity of H_2O molecules. Likewise, people throughout history have been aware of lightning, but many of them would be quite surprised to find out that it is an electric discharge. Once again, conceptual distinction doesn't make for actual distinction, and empirical investigation can reveal to us that an identity actually underlies a conceptual distinction.

The identity theorist will begin by allowing for a *conceptual distinction* between "Patrick sees red" and "A c-neural network is active in Patrick's brain." These two phrases do not mean the same thing, and somebody could be aware of the one while even doubting the existence of the other. Nevertheless, some materialists claim that (24) is an empirical identity:

(24) Patrick's seeing red is a c-neural network activation in Patrick's brain.

As J. J. C. Smart puts it, "Sensations are nothing over and above brain processes. Nations are nothing 'over and above' citizens, but this does not prevent the logic of nation statements being very different from the logic of citizen statements."[27] In other words, Patrick's seeing red *just is* the activation of his c-neural network, in the same way that water just is H_2O and lightning just is electric discharge; in all of these cases a conceptual distinction hides a numeric identity revealed by scientific investigation. That is, identity theorists will grant Descartes his basic claims about the conceivability of mind without body, but show that this doesn't make for actual distinction between entities. It is conceivable for

27. Smart, "Sensations and Brain Processes," 56–57.

there to be lightning without electric charge, since I could be acquainted with lightning without being acquainted with electric charge. If we were to travel in time back to the thirteenth century and offer a leading scientist of that era (though he would call himself a *natural philosopher*) a description of electrical charges, he would find it entirely conceivable that lightning exists without electric charge. Nevertheless, subsequent scientific investigation has revealed that there is an electric charge–lightning identity relation. Likewise, sensations and thoughts *seem* very different from the activation of neural networks, but the same can be said about water and H_2O. It might be unthinkable to a scientifically primitive person that water would actually be a collection of particles, just as our medieval scientist would doubt the identity of electric charge and lightning. In all of these cases, however, empirical investigation has revealed to us that there is an underlying identity. As Searle puts it, "Descartes might have been right in thinking that there were separate mental phenomena; it turned out as a matter of fact that he was wrong," and for this reason we shouldn't give any credence to the modal argument.[28]

Moreover, if we have good reason to believe that propositions like (24) are empirical identities, we would then have good reason to doubt the crucial inference presupposed by certain versions the difference argument, which you will remember is the strongest part of the case for dualism.[29] The difference argument presupposes the following inference:

(25) The mind has a property, *F,* and the brain lacks *F.*

Therefore:

(26) The mind and the brain are not identical.

If we accept that psychological states might be identical to brain states, even though they are conceptually distinct, we have very

28. Searle, *The Rediscovery of the Mind,* 36.

29. These considerations would seem to do little, however, to undermine the unity of consciousness argument (advanced by Hasker) or the argument from the problem of the many (advanced by Unger). See chapter 2, "The Difference Argument for Dualism."

good reason to doubt (25), and therefore the inference to (26) would not stand. What seems like a real difference, even an essential difference, according to our a priori inspection of our psychological states, might turn out, after empirical investigation, to be just two different appearances or presentations of the same self-identical thing. Thus, we might find out that a sensation or thought turn out to be one and the same thing as a certain neurophysiological state. The Morning Star and the Evening Star certainly *seem* to be different objects, but they are not, just as lightning and water might *seem* to be essentially very different from electrical discharge and H_2O respectively, though they are not. Likewise, argues the identity theorist, sensations and thoughts seem to be essentially different from brain states, but we have empirically discovered that they are not.

Has neuroscience revealed an empirical identity between psychological states and brain states? It clearly has not, nor does it seem that it ever could. What neuroscience does is to provide us with ever more precise descriptions of the goings-on in our brains that are *correlated* with our psychological states. The neuroscientist can, within a certain range of specificity, show us that when Patrick sees red, his c-neural network is active. Correlation is a far cry from identity; two phenomena can be perfectly correlated, without being identical. Correlation is a good first step toward demonstrating causation and even identity, but it does not constitute a case by itself. J. J. C. Smart argues, however, that the case for the identity theory rests not on the direct empirical evidence, but on theoretical simplicity: "If it be agreed that there are no cogent philosophical arguments which force us into accepting dualism, and if the [identity theory] and dualism are equally consistent with the facts, then the principles of parsimony and simplicity seem to me to decide overwhelmingly in favour of the [identity] theory."[30] In other words, if we assume that dualism is not demonstrably certain and

30. Smart, "Sensations and Brain Processes," 66.

that both identity theory and dualism can account for the relevant facts, then we should prefer the identity theory because it asks us to believe only in entities we can empirically verify. Let's suppose then that dualism is not demonstrably certain, so then our next question will be whether identity theory can account for all of the facts regarding our psychological lives, and this is an issue we will consider for the remainder of this chapter.[31]

Before we consider how identity theorists grapple with the psychological facts, we should pause to address one fairly obvious objection. Searle summarizes the objection succinctly:

> Either the pain features are subjective, mental, introspective features, or they are not. Well if they are, then we have not really gotten rid of the mind. We are still left with a form of dualism, albeit property dualism rather than substance dualism. We are still left with mental properties, even though we have gotten rid of mental substances. If on the other hand we try to treat 'pain' as not naming a subjective mental feature of certain neurophysiological events, then its meaning is left totally mysterious and unexplained. As with behaviorism, we have left out the mind. For we now have no way to specify these subjective mental features of our experiences.[32]

In other words, even though we grant that "The Morning Star is the Evening Star" states an empirical identity, *they* are both Venus; we do not deny the reality of the properties *being the Morning Star* and *being the Evening Star.* The point is that empirical identity doesn't make real distinctions between properties go away. Now, consider once again

(27) Patrick's seeing red is a c-neural network activation in Patrick's brain.

This identity doesn't tell us that there is no such thing as *seeing red,* just that *seeing red* is a property of something that also has the property of *being a c-neural network activation.* Though prop-

31. Other identity theorists have argued for their positions in various other ways, including appeals to causal closure, parsimony, and broader philosophical commitments to physicalism. See Polger, "The Identity Theory," 829–30, and Kim, *The Philosophy of Mind,* 98–121.

32. Searle, *The Rediscovery of the Mind,* 37.

erty dualism might be preferable for the naturalist, it still raises many of the problems that the materialist has cited as reasons for rejecting substance dualism; for example, the property dualist would still need to explain how nonphysical properties can interact with physical events in the brain. The identity theorist may attempt to deny the reality of mental properties, but this would seem to invoke eliminative materialism and all of its difficulties.[33]

J. J. C. Smart suggests a reply to the property dualism objection in his classic paper, "Sensations and Brian Processes," by introducing *topically neutral definitions* of psychological states. Consider the following scenario.[34] William answers the telephone and the caller identifies himself as the dentist. Notice that simply in virtue of taking the call William is acquainted with an intrinsic property of the caller, that is, his being the dentist. Thus, William's acquaintance with the caller is not entirely neutral to the caller's intrinsic nature. Should William later learn that "the dentist is the philatelist" is a true identity statement, he would have to conclude that the caller has two different properties, *being the dentist* and *being the philatelist.* Now suppose that William answers the telephone, and the caller does not identify herself. William's acquaintance with the caller is neutral with respect to her intrinsic nature, such that William would be able to give only a neutral description like the following:

(28) Somebody called.

Notice that (28) is indifferent to who the caller is, and there is nothing with which William is acquainted that gives him any hint as to the intrinsic nature of the caller.[35] Now, suppose there are

33. Some eliminative materialists take such an extreme position in order to avoid the property dualism they see inherent in standard versions of the identity theory. See Rorty, "Mind-Body Identity, Privacy, and Categories," and Feyerabend, "Mental Events and the Brain."

34. In what follows I am varying an example that Armstrong introduces in *The Mind-Body Problem: An Opinionated Introduction.*

35. For the sake of the simplicity of the example, leave aside properties such as the caller's tone of voice or trivial properties like his or her being a caller, etc.

two competing hypotheses about the intrinsic nature of the caller; that is, the caller is the dentist or the caller is the veterinarian. Thus, William is considering adopting one of the two following identifications of the caller:

(29) The dentist called.

(30) The veterinarian called.

Notice that (29) and (30) are not neutral as to the identity or nature of the caller. Suppose that eventually William adopts, for say empirical reasons such as "Today is Friday, and the veterinarian never calls on Friday," a theory such that it leads him to conclude:

(31) The dentist called.

Does this mean that William must say that that the caller has two sets of properties, that is, *being the dentist* and *being the veterinarian*? No! He has discovered that the previously unknown *somebody* who called is in fact the dentist. It is not as though he was directly acquainted with the caller as both the dentist and the veterinarian. Rather, his initial encounter with the caller was neutral with respect to these descriptions, but he later discovers empirically that the caller is the dentist. There is no reason whatsoever to attribute *being a veterinarian* to the caller.

Smart argues that we can give topically neutral definitions of psychological states: "My suggestion is as follows. When a person says, 'I see a yellowish-orange after-image,' he is saying something like this: 'There is something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange.'"[36] Smart's point is that we can give a topically neutral definition of the sensation of orange (or any sensation for that matter) like the following:

36. Smart, "Sensations and Brain Processes," 61. Place makes a similar point in accusing the dualist of falling for what he calls the phenomenological fallacy; see Place, "Is Consciousness a Brain Process?" 58–60.

(32) The sensation of orange is the state that results when someone with functioning senses is in the presence of an orange object under standard lighting conditions, and so on.

Thus, when we say, for example, that

(33) William is seeing orange

what we have really given is a topic neutral description of Will's state:

(34) William is in a state that results when someone with functioning senses is in the presence of an orange object under standard lighting conditions, and so on.

Notice that (34) is neutral between

(35) William is in a private, incorrigible, nonphysical state.

and

(36) William's d-neural network is activated.

If (33) is really a topic neutral report, best understood as (34), then whether we commit ourselves to (35) or (36) is a matter not of direct acquaintance, but of broader theoretical and empirical considerations. Since we have defined "sensation of orange" in topically neutral terms, should we adopt (36), there is no lingering property dualism, because neither (35) nor (36) is entailed by our topically neutral description. If (36) is vindicated theoretically or empirically, there is one property, for example, d-neural network activity, that we are ascribing to William, which, it turns out, just is what it is to have a sensation of orange.

How does topic neutrality help the identity theorist? Smart's point is that the identification of psychological states with neurophysiological activity does not require that we recognize both nonphysical and physical properties. Rather, the nonphysical and the physical descriptions are competing theoretical accounts of the topic neutral states with which we are acquainted directly. On Smart's rendering, the identity theory shows not that we have one state with two different sets of properties, one mental and one physical, but that what we previously described in topically

neutral terms can now be given a more theoretically rich neurophysiological description. As Armstrong puts it, "Our mind is not experienced by us as immaterial or material. What its true nature is, is a matter of theory."[37] Once we accept the default topic neutrality of our experience of the mind, the only question is which description is theoretically preferable, and at this point the "scientific considerations that favor the materialist hypothesis are then appealed to in deciding the issue between the rivals."[38] Given her broader commitments to the failures of dualism and the theoretical virtues of materialism, the identity theorist believes the neurophysiological description has the advantage.

Be careful not to mistake identity theory for eliminativism, although the distinction between the two is subtle. *The identity theorist, contrary to the eliminativist, does not deny that there are psychological states.* There are, according to the identity theorist, such events as sensations and thoughts. What the identity theorist does deny is that sensations and thoughts are mental or nonphysical. On this view, empirical investigation, along with a bit of theorizing, has revealed that Patrick's *tasting an orange* and *thinking that "9 x 9 = 81"* are neurophysiological states. We need not deny that Patrick thinks and feels, so long as we understand that thinking and feeling really are just physical processes.

Smart's topic neutrality solution to the problem of property dualism is ingenious, no doubt. It is, however, the subject of serious objections. Since topic neutrality is the cornerstone of most subsequent developments of identity theory, I want to leave these problems aside for the moment, so that we may have the most sophisticated version of materialism in mind before we consider criticisms. We will spend the entire next chapter evaluating materialism in light of these objections.

37. Armstrong, *The Mind-Body Problem*, 79.
38. Ibid., 78.

Token Identity Theory and Nonreductive Materialism

For now, we will consider a series of objections that have challenged materialists to develop a more refined version of identity theory. Before we can understand this fascinating line of philosophical development, we must first get a hold on a pretty easy distinction between *type* and *token.* By a *type* we mean a *kind* or *sort* of thing—what is more traditionally, and more controversially, called a *universal.* A token is an *instance* or *particular* of a type; what is frequently called an *individual.* For example, "r" and "*ℛ*" are two tokens of the type letter "R"; Martha and Patrick are both tokens of the type human being; the chair I am sitting on is a token of the type chair.

Identity theory as we have considered it so far may be interpreted as asserting a *type-identity* relation between the psychological states and neurophysiological states. That is, the identity theorist might argue that every type of psychological state is identical to a single type of neurophysiological state; the identity obtains between two types of things, a psychological state type and a neurophysiological state type. Take the psychological state type pain and a hypothetical neurophysiological state type, c-neural network activation. The type-identity version of materialism would then claim:

(37) Pain is c-neural network activation.

On this view, there is no difference between having a pain and having c-neural network activation; they are the same type of event. Of course, type-identity implies something about tokens of those types, that is, every token of pain is identical to some token of c-neural network activation, such that we may define *type-materialism* as follows: For every psychological state type, *P,* there is a single neurophysiological state type, *B,* such that every token of *P* is a token of *B* and every token of *B* is a token of *P.*

Ned Block objects to type-identity theory, because such a

theory "is a chauvinist theory; it withholds mental properties from systems that in fact have them. In saying mental states are brain states, for example, physicalists unfairly exclude those poor brainless creatures who none the less have minds."[39] You, like some of my students, may be puzzled by what seems like Block's charge of sexism against type-identity theory. Is this a case of hyperbolic political correctness gone awry? Actually, Block is using the term "chauvinist" in its broader sense, meaning an unfair devotion to any particular group at the expense of others. Block's point is that type-identity theory would exclude the possibility of psychological beings that don't have any neurophysiology whatsoever, such as "computing machines, any pneumatic or gaseous extraterrestrial beings we may encounter, or wooden puppets that have been swallowed by whales."[40] Even if we are not impressed with the possibility of such non-neurophysiological beings having psychological states (though in the following section we will take some of these proposals rather seriously), there are certainly many other organisms with neurophysiologies that we want to say are subject to pain (among other psychological states) and yet have neurophysiologies quite different from our own. The point is that psychological states are capable of *multiple realizations* in very different neurophysiologies.[41] The problem is that if we accept (37) as our account of pain, then any organism that doesn't have exactly c-neural networks doesn't have pain. If we take type-identity theory seriously, then to have a pain *is just* to have a c-neural network activation, and if an organism lacks that type of neural network, then it lacks pain. Indeed, it is very difficult to discern exactly what makes two tokens members of the same neural network type, since no two organisms, even or-

39. Ned Block, "Troubles with Functionalism," in *Mind and Cognition: A Reader,* ed. William Lycan (Oxford: Routledge, 1990), 446.

40. Polger, "Identity Theories," 826.

41. This argument is classically advanced by Hilary Putnam, "Psychological Predicates," in *Art, Mind, and Nature,* ed. W. Captain and D. Merrill, 229–40 (Pittsburgh, Pa.: University of Pittsburgh Press, 1967).

ganisms of the same species, have brains that are wired exactly alike. It seems that if we pushed this objection to its fullest extent, the type-identity theory would be so chauvinistic as to deny pain to all organisms, but a single individual. This is clearly an absurd consequence.

A promising reply to this objection is to amend materialism in terms of *token-identity*. Whereas type-identity identifies types, token-identity does not involve mapping one type onto another. Rather, in token-identity the identity relation obtains only at the level of tokens, such that all tokens of some type are identical to tokens of a certain *open-ended range* of other types. For example, suppose that we know:

(38) Every member of the Madden family is a member of a sports team.

We could interpret (38) in terms of type-identity, which would mean that for the type *Madden family,* there is a single sports team, for example, the wrestling team, to which all the Maddens belong. If that interpretation were accused of athletic chauvinism, we could then offer a token-identity interpretation:

(39) Every member of the Madden family is a member of *some sports team or other,* that is, the wrestling team *or* the football team *or* the bowling team *or* the track team, and so on.

Our thesis now is only that all of the Maddens are members of some team, but there is not necessarily any particular team to which they all belong, nor do we assume a final list of team types could be assembled. The charge of athletic chauvinism is thereby averted. Searle discusses the application of token-identity in these remarks:

In any case, the idea that any type of mental state is identical with some type of neurophysiological state seemed really much too strong. But it seemed that the underlying philosophical motivation of materialism could be preserved with a much weaker thesis, the thesis that for every token instance of a mental state, there will be some token neurophysiological event with which that token instance is identical. Such views were

called "token-token identity theories" and they soon replaced type-type identity theories.[42]

Materialism construed in terms of type-identity claims that all tokens of a certain psychological state type are identical to tokens of one neurophysiological state type; there is a one-to-one correspondence between psychological and brain state types. As Searle suggests, materialism construed in terms of token-identity does not claim that there is a one-to-one correspondence between psychological state types and brain state types, only that each token of a psychological state is identical to a token of *some brain state type or the other;* the identity occurs only at the level of tokens, not types.[43] We may define token-token identity materialism as follows: *For every token of a certain type of psychological state, there is some token of some type of physical state or other to which that psychological state token is identical.*

The token-token version of materialism solves the problem of neural chauvinism. Now the materialist need not claim, as we say in (37), that pain is c-neural network activation, thereby denying pain to organisms lacking c-neural networks. Rather, the token materialist may offer the following account of pain:

(40) Any token of pain is a token of c-neural network activation *or* d-neural network activation *or* e-neural network activation, *or* ....

Thus, we have not identified pain with a single brain state type, but with an open range of state types. It's important to emphasize that this range of physical state-types to which a psychological state might be token-identical is indefinite; that is, there is no reason to suppose that we have (or even could have) a complete list of all of these physical state types. Thus, we may be more inclusive of organisms that come in many neurological varieties, or, to

42. Searle, *The Rediscovery of the Mind,* 39–41.

43. For an example of token-physicalism, see Colin McGinn, "Anomalous Monism and Kripke's Cartesian Intuitions," *Analysis* 37, no. 2 (1977): 78–80; and Noah Latham, "What Is Token Physicalism?" *Pacific Philosophical Quarterly,* 84, no. 3 (2003): 270–90.

take Block's example, even those that seem to lack sophisticated neurology at all.[44]

The term "reduction" is not philosophically innocent, but if we take "reductive materialism" to mean that "minds are 'nothing over and above' brains or physical stuff," then the type-identity theory, that is, the claim that there is a type-identity between psychological and neurophysiological states, is the paradigmatic case of reductive materialism.[45] Notice, however, that in this sense we should likewise construe token-identity theories as reductive, because these views (despite their differences) agree with the type-identity theory that instances of thoughts and sensations involve "nothing over and above the physical/brain stuff."[46] That is, all of these views (at least as employed by materialists) are reductive in the sense that when an organism (or possibly a machine) has a token of a sensation or a thought, that sensation or thought just is a token of a physical state (whether it be a state of a nervous system or not).

There is another sense of "reductive" in which token-identity theories can be fairly said to be nonreductive in contrast to the type-identity theory. If the type-identity theory is true, then we should be able to reduce our psychological theories to physical theories, in particular neurological theories, in the same way that in physics we can reduce thermodynamics to statistical mechanics or optics to electromagnetic theory. In these examples, we are able to align the laws of thermodynamics and optics with those of statistical mechanics and electromagnetic theory respectively in a way that shows that thermodynamic phenomena and optical phenomena, though they are real as opposed to illusory, actually

44. For further development of the token-identity approach, see Jaegwon Kim, "On the Psycho-Physical Identity Theory," *American Philosophical Quarterly* 3 (1966): 227–35; and Donald Davidson, "Mental Events," in *Essays on Actions and Events,* 207–25 (Oxford: Clarendon Press, 1980).

45. Polger, "Identity Theories," 825. Polger offers a helpful discussion of several senses in which identity theories are or are not reductive in this article.

46. Ibid.

just are the phenomena of statistical mechanics and electromagnetic theory. In other words, we can map the one *theory* onto a broader theory in way that allows us to conclude, for example, that temperature just is the mean kinetic energy of molecules, and so on. The type-identity theory would seem to involve just such a model of inter-theoretical reduction:

> Mind-body reduction, on this model, would consist in the derivation of psychological laws form those of some underlying physical theory (presumably, neurobiology) taken together with "bridge laws," empirical laws correlating mental kinds with physical-neural kinds. The idea was the reduction could be carried out by finding for each mental property a nomologically coextensive physical property.… In any case, the idea that mind-body reduction must be underwritten by a pervasive system of type-type correlations has been the tacit but widely shared presupposition of the recent debate on reductionism.[47]

As we emphasized above, according to token-identity theories, however, there is no way we can complete the sentence "To have a sensation/thought of type M, is to be in physical state _____," because tokens of an indefinite number of physical state types could be tokens of a psychological state type. Thus, token-identity theorists grant that there is no way for us neatly to map our ordinary idiom of folk psychology onto any physical idiom or theory, even though they claim that it so happens that all psychological state tokens are in fact tokens of physical types. As it is sometimes put, there are no psycho-physical bridge laws by which an inter-theoretic reduction could be executed.[48]

Donald Davidson develops a highly influential theory that is both a version of token-identity theory and nonreductive in this sense.[49] Davidson argues that since psychological states are

47. Jaegwon Kim, "Supervenience," in *A Companion to the Philosophy of Mind,* ed. Samuel Guttenplan, (Cambridge, Mass.: Blackwell, 1995), 579. See also, Jaegwon Kim, "On the Psycho-Physical Identity Theory." Note that Kim is using the term "mental" in the sense that we have been using "psychological" throughout this book.

48. Polger argues that there are some versions of the type-identity theory that are not reductive in the inter-theoretic sense; see "Identity Theories," 825. Thanks to an anonymous reviewer for leading me to clarifying the notion of reduction.

49. Though Davidson's position is influential, not all nonreductive materialists hold

anomalous in the sense that since there are no "strict laws" that explain or predict them, it follows that no "mental phenomena can be given purely physical explanations," because physical explanations involve strict laws.[50] This position, which Davidson calls *anomalous monism,* is a nonreductive materialism, because it recognizes both the explanatory autonomy of psychological states and the token-identity of those states with physical states; psychological beings are thoroughly physical, even though their psychological states cannot be explained in the same terms as physical states. Davidson affirms both that "mental events are, in my view, physical (which is not, of course, to say that they are not mental)" and "There are no strict psychological laws (laws connecting mental events under their mental descriptions with physical events under their physical descriptions)."[51]

Questions abound regarding nonreductive materialism, but the most pressing regards exactly how we are to distinguish nonreductive materialism from property dualism. The nonreductive materialist claims that when an organism has a sensation or thought it is in an entirely physical state, but there is supposed to be no way we can identify *feeling pain* or *thinking of the greatest prime number* with any physical state, say *c-neural net activation.* As Andrew Melnyk, a critic of nonreductive materialism, puts it: "The rub, however, is the event-identity [token-identity] claim tells us nothing about the relations between the single event's having the mental property that it has and its having the physical property it has; a fortiori, the event-identity [token-

this particular theory. We will consider a number of variations on this theme in our subsequent discussion.

50. Davidson, "Mental Events,"212, 214.

51. Donald Davidson, "Davidson, Donald," in *A Companion to the Philosophy of Mind,* 31. Note that in this passage Davidson uses "mental" in the same sense as we have used the term "psychological" throughout this book. We will consider criticism of nonreductive materialism in general in the next chapter too, but for an introduction to criticism of Davidson's anomalous monism in particular, see Andrew Melnyk, "Can Physicalism Be Non-Reductive?" *Philosophy Compass* 3/6 (2008): 1281–96. Melnyk's paper also contains a fair (though critical) survey of various nonreductive materialist positions.

identity] claim yields no sense in which the property that it has is *nothing over and above* its having the physical property that it has."[52] What is the relation between these psychological states and physical states such that what we have in the nonreductive materialist account is not a form of property dualism? The most prominent answer to this question, first recommended by Davidson and extensively developed by other thinkers, is couched in terms of *supervenience.*[53] Suppose Martha sculpts a statue of Albus Dumbledore, and in doing so she forms the clay into a particular configuration, *C1.* We cannot strictly reduce *being a statue of Dumbledore* to *being clay in C1,* because indefinitely many other configurations are likewise sufficient for *being a statue of Dumbledore.* Notice, however, that *being clay in C1* necessitates *being a statue of Dumbledore* (the former alone is sufficient for the latter). Thus, this statue could not fail to be a statue of Dumbledore without failing to be the clay in *C1.* Finally, though *being a statue of Dumbledore* is irreducible (in the strict sense) to *being clay in C1,* this token of *being a statue of Dumbledore* is nothing over and above, to use Melnyk's phrase, this token of *being clay in C1;* there isn't any property had by this statue that isn't really just a property of this clay so-configured. The relationship between this token of *being a statue of Dumbledore* and this token *of being clay in C1* is what we mean by supervenience; this token of *being a statue of Dumbledore* supervenes on this token of *being clay in C1.*

One way that philosophers have commonly specified *psychophysical supervenience* is as follows: *If something has a psychological property, M, then that thing has a physical property, P, and necessarily anything that has P has M.*[54] Just like our example of statue-

52. Melnyk, "Can Physicalism Be Non-reductive?" *Philosophy Compass* 3, no. 6 (2008): 1285. Author's emphasis.

53. What follows is a very simplified discussion of a highly complex topic in the philosophy of mind, which has generated a legion of literature. The most reliable introduction is in Kim, *The Philosophy of Mind,* 8–14 and 122–27. See also Kim's "Supervenience" and *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993).

54. I've adopted this way of construing psycho-physical supervenience from what

configuration supervenience above, psycho-physical supervenience entails radical dependence of the psychological on the physical; that is, there can be no variation in the psychological without a variation in the physical, even though it does not entail a strict reducibility of psychological states to physical states. Contrary to what we might expect from a property dualist, the token-identity theorist can claim that psychological states are not free-riders; that is, they are firmly based on or grounded in the physical states on which they supervene. Moreover, if psycho-physical supervenience is analogous to statue-configuration supervenience, then it does seem that being a token of a psychological property is nothing more than being a token of any one of an indefinite range of physical types, just as the nonreductive materialist would have us think.

Functionalism and Artificial Intelligence

The appeal to psycho-physical supervenience has every appearance of being promising, but it likewise creates some work for the materialist. Terence Horgan points out that what the nonreductive materialist needs is not "a bare supervenience" but "*superdupervenience:* viz., ontological supervenience that is robustly explainable in a materialistically explainable way … which thereby confers materialistic respectability."[55] We can readily "see" how the configuration necessitates the statue of Dumbledore simply in virtue of its properties as a certain configuration, such that in this case the supervenience relation is transparent or at least readily explicable. Therefore we should likewise expect a similar sort of explanatory transparency in the relation between psychological states and their supposed physical-supervenience bases. In other words, we should be able to understand why a

Kim calls "strong supervenience," and do note that there are other versions. See *The Philosophy of Mind,* 9–10, and *Physicalism or Something Near Enough,* 32–38.

55. Terence Horgan, "From Supervenience to Superdupervenience," in *The Philosophy of Mind: Classical and Contemporary Readings,* ed. David Chalmers (New York: Oxford University Press, 2002), 154. Author's emphasis.

certain token of a physical state necessitates a token of a psychological state simply in virtue of its physical properties. In short, the challenge for the nonreductive materialist is to provide a satisfying account of psycho-physical supervenience that makes sense within the parameters of materialism more generally.

Thus, when the nonreductive materialist claims that, for example, tokens of pain supervene on tokens of an open-ended range of neurophysiological state types, it is fair to wonder what all of these neurophysiological state types have in common such that pains supervene on them. The type-identity theorist doesn't really have a problem here, as he can claim that pain *just is* a certain neurophysiological state. The nonreductive materialist, however, grants that psychological state types cannot be identified with any physical state types, so it makes sense to wonder why psychological states happen to supervene on any particular physical state. What we are looking for is a definition or characterization of psychological state types (what Ned Block calls a *principle of individuation* for psychological states) that sheds light on why they supervene on certain physical states.[56] For example, in the case of (40) above, what do type c-neural network activation, type d-neural network activation, and type e-neural network activation all have in common such that tokens of pain supervene on them? In short, what we are asking from the nonreductive materialist is an account of psychological states that explains why they happen to supervene on certain physical states, but not others.

The most influential answer to this problem is *functionalism.*[57]

56. See Ned Block, "Introduction: What Is Functionalism?" in *Readings in Philosophy of Psychology*, vol. 1 (Cambridge, Mass.: Harvard University Press, 1980), 178–79, cited in Kim, *The Philosophy of Mind,* 102.

57. Jerry Fodor, "Materialism," in *Materialism and the Mind Body Problem,* 2nd ed., ed. David Rosenthal (Indianapolis: Hackett Publishing, 2000), 128–49. Other early, innovative, and influential versions of functionalism are developed by Hillary Putnam, "Psychological Predicates"; David Lewis, "Psychophysical and Theoretical Explanations," *Australasian Journal of Philosophy* 50, no. 3 (1972): 249–58; and David Armstrong, *A Materialist Theory of Mind* (London: Routledge and Kegan Paul, 1968). Lewis's and Armstrong's view's are variously interpreted as a versions of identity theory or functionalism.

As Jerry Fodor points out, functional analysis defines "a part of a mechanism" by "what role it plays in the activities that are characteristic of the mechanism as a whole: 'What does the camshaft do?' 'It opens the valves, permitting the entry into the cylinder of fuel.'"[58] We define a camshaft in terms of a functional analysis, because a camshaft just is whatever performs the valve-opening role in a combustion engine. We can be completely neutral about the physical embodiment of camshafts: it doesn't matter whether the part is steel, plastic, porcelain, wood, or any other material; as long as it plays the proper role in the engine, it's a camshaft. The nature of a camshaft is defined by its performance of a certain function, not its composition by a certain kind of stuff. The functionalist argues that tokens of various neurophysiological states can be classified as supervening on tokens of a certain psychological state type not because of their shared intrinsic properties, but because of their common function or role in the organism. That is, for the functionalist psychological state types are defined in terms of roles in the overall behavior of an organism independently from their particular realization in tokens of neurophysiological states of such organisms. Thus, any token of a neurophysiological state that plays a certain role is the supervenience base (or realizer) of a token of a certain psychological state type.

To see how this works, consider:

(41) Brendan believes that the stove is hot.

A functionalist would begin her analysis by claiming that Brendan's believing is token-identical to a certain functionally defined state such that:

(42) Brendan is in an r-neural network activation caused by the heat stimulus from the stove that along with his desire not to be burned in turn caused the retraction of Brendan's hand.

Notice, the claim here is that having the belief that the stove is hot just is to have a state that is caused in a certain way and causes

58. Fodor, "Materialism," 144.

certain other states or activities. The psychological state is whatever plays the intermediary role between the stimulus and the response, between the causal input from the environment and the output of behavior. Any internal state that plays this causal role in an organism is a belief that the stove is hot—no matter what its actual embodiment might be. As such consider the following:

(43) Kermit (a frog) believes that the stove is hot.

We can analyze (43) as:

(44) Kermit is in an f-neural network activation caused by heat stimulus from the stove that along with his desire not to be burned in turn caused the retraction of Kermit's limb.

We can say that Brendan and Kermit both had the same belief even though their beliefs supervene on tokens of different brain state types, because those brain state tokens are functionally equivalent, that is, they play the same role in their respective behaviors.

Of course, a worry looms here because both (43) and (44) contain some old-fashioned psychological language referring to Brendan and Kermit's shared *desire* not to be burned. This problem is fixed easily enough if we assume that desires can be functionally specified in terms of roles neurophysiological states play in linking stimuli and responses. Assuming we have such an analysis of the desire not to be burned, our final analysis of Brendan's belief that the stove is hot would be:

(45) Brendan's state of being in a state of q-neural network activation caused by heat stimulus from the stove caused him to be in a state of r-neural activation that along with his prior state of t-neural activation caused his hand to recoil.

It seems perfectly plausible that we would give such an analysis of the various neurological states that link stimuli to responses in Brendan. There likewise seems to be nothing barring us from specifying neurophysiological states that play much the same role in the stimulus-response loop in Kermit. Thus, it seems perfectly plausible that a characterization of psychological states in terms of functional roles can be given, which is consistent with

such states being identical to tokens of various neurological state types. We then need not posit anything nonphysical to make sense of our behavior. We need, however, to be careful not to confuse functionalism with behaviorism. The functionalist claims not that psychological states are behaviors, but that psychological state types can be defined in terms of functions linking environmental inputs with behavioral outputs. When an organism is stimulated in some way, that input is somehow converted into a behavioral output.[59] Whatever does this behind-the-scenes work is the psychological state. In short, *the psychological state is not the behavior but something that plays a role internal to the organism linking stimuli to behavioral responses.* Thus, the functionalist avoids the absurdities of behaviorism; that is, functionalism preserves psycho-physical causation while avoiding the problems of circularity and good liars, but exploits its root insight: overt behavior does play a role in how we ascribe mental states to organisms.[60]

It should be clear that functionalism can go a long way toward showing how the supervenience relation between a psychological state and its physical base can be explained. A successful functionalist program would seem to answer this concern, because "if mental properties as a family were identical with certain function-

59. For the sake of simplicity, I am leaving aside cases in which the internal state causes another internal state in reaction to an external input. It is easy enough to see how the functionalist can make sense of these cases too.

60. I have tried to present as generic a version of functionalism as possible, though it actually comes in significantly different varieties. These distinctions are important, but they require too much in the way of technical detail to be helpful to the novice reader. For a discussion of different versions of functionalism, see Armstrong, *The Mind-Body Problem,* 105–10, and Kim, *Philosophy of Mind,* chapters 5 and 6. For examples of *causal role functionalism,* see Armstrong, *A Materialist Theory of Mind,* and Lewis, "Psychophysical and Theoretical Explanations." For an example of *teleological functionalism,* see Elliot Sober, "Putting the Function Back into Functionalism, in *Mind and Cognition,* 97–106. For an example of *humuncular functionalism,* see Daniel Dennett, "Why the Law of Effect Will Not Go Away," in *Mind and Cognition,* 63–73. For an example of *evolutionary functionalism,* see Ruth Millikan, *Language, Thought, and Other Biological Categories: New Foundations for Realism* (Cambridge, Mass.: MIT Press, 1984). We will discuss specifically teleological and evolutionary versions of functionalism in the following chapter in response to objections.

al properties whose definitive causal roles involve typical-causal relations to sensory stimulation and bodily motion, and one another, then specific physical/mental supervenience relations presumably would be materialistically explainable in terms of causal/dispositional roles of categorical physical properties."[61] If we ask, "Why does this token of particular psychological state *M* supervene on a token of physical state *P*?" the functionalist has a ready answer: "Because *P* realizes the causal-functional role that defines *M*." Moreover, we can in principle specify the causal-functional role that defines *M* in terms of stimulus response relations that involve nothing more than straightforward physical causation. Functionalism then seems to be just what the doctor ordered for nonreductive materialism, as it is nonreductive in the strict sense (we cannot identify any psychological types with physical types) and it provides an explanatory and materialist-friendly characterization of supervenience: psychological states supervene on the physical states that happen to realize them, and why this is the case can be easily explained in terms of the casual-functional roles these physical states play.[62]

We must be careful, however, when we mention functionalism in the same breath as nonreductive materialism, because functionalism need not be a version of materialism at all! You may have noticed that we have made mention only of neurophysiological states in our functionalist analyses so far, but this need not be the case. Remember that our functional analysis of the nature of a camshaft was topic neutral, and therefore presupposed no particular material composition; that is, the camshaft can be made of vanilla pudding as well as steel, so long as it plays the

61. Horgan, "From Supervenience to Superdupervenience," 158.

62. I have been using the term "realize" alongside "supervene" in this section, but some care should be taken here, as philosophers have construed "realization" as a stronger relation than "supervenience." Nevertheless, I'm following Kim ("Supervenience," 581) when he says, "Those who use the idiom of 'realization' will agree that physical duplicates realize exactly the same set of mental properties. . . . The idea that mentality is physically realized is integral to the functionalist conception of mentality, and this commits most functionalists to mind-body supervenience in one form or another."

proper functional role. The functional analysis of psychological states is no less ontologically neutral; it doesn't matter what particular ontological mode it takes, so long as it plays the appropriate linking role between stimulus and response. The functional state could be instantiated in an animal's nervous system, or it might not even be instantiated in an organism. Indeed, there is nothing about functionalism as such that requires that functionally characterized, psychological states be realized physically. Even though most functionalists would say that it so happens that psychological states in the actual world are realized only in physical systems, that is something accidental to functionalism per se. The success of functionalism is, however, very good news for the nonreductive materialist, as we discussed above, because it purports to show that psychological states can be realized by completely physical systems, even if there could be nonphysical systems that realize psychological states. Since, as the functionalist claims, psychological states don't entail anything about the physical or nonphysical status of their bearers, we are then free to embrace materialism for empirical reasons, for example, the results of neuroscience, or broader metaphysical concerns, for example, causal closure. Thus, even though functionalism is not strictly a materialist theory (as it is metaphysically neutral), its success as an account of psychological states would be so helpful to the general case for materialism that in what follows we will treat it as the most promising heir to the materialist's program. As Suzanne Cunningham puts it, "while early Functionalists said that a system that was capable of the relevant functions could be made of anything, including nonphysical substance, the vast majority of Functionalists are committed to the view that the relevant systems are indeed physical."[63]

Even leaving aside the possibility of nonphysical realizers of psychological states, there is no reason for the functionalist to

63. Suzanne Cunningham, *What Is a Mind?* (Indianapolis, Ind.: Hackett Publishing, 2000): 48. Thanks to an anonymous reviewer for pressing me to clarify this point.

limit the realization of psychological states to biological systems. For example, let's define the feeling of heat as follows:

(46) Feeling heat is any token of a type-t functional state caused by heat input that causes a token of s-type behavior as an output.

Any entity that has a state that satisfies (46) would then be said to feel heat, even if such an entity is a machine—feeling heat just is being in such a functional state. Thus, *for some strict functionalists, the thermostat in your living room, though in a very primitive way, feels heat; that is, it is a conscious entity.*

The functional state could be embodied in a silicon, plastic, or metal system, for example, a digital computer, so long as it instantiates the appropriate functional role. Those who take (46) as a sound account tend to accept what Searle calls *strong artificial intelligence* (AI): appropriately programmed digital computers have minds in the same sense that human beings have minds.[64] Why does functionalism lead one to strong AI? A digital computer just is a system (hardware) such that any well-formed, particular input into the system leads to a definite output as determined by the particular software the system has been configured to run. Thus, the activation of the software is the causal link between the inputs into the system and its outputs. As long as these inputs and outputs can be defined in terms of perceptual stimulation and behavioral responses, it would seem that functionalism implies that software is a type of mind. As the typical analogy goes, mind is to brain as software is to hardware. If functionalism is our best theory of mind, then it seems perfectly plausible that highly sophisticated machines would have beliefs and sensations, because such psychological states are defined in terms of functional roles, and nothing bars a machine from embodying the appropriate functions.[65]

64. See Searle, *The Rediscovery of the Mind,* 43.

65. See Eric Steinihard, *More Precisely: The Math You Need to Do Philosophy* (Peterboro, Ontario: Broadview Press, 2009), 65–85, for a good introduction to the mathematical and logical background to the philosophical discussion of functionalism and artificial intelligence.

The question arises as to how we would know whether a nonbiological system is actually conscious. As John Haugeland points out, once we have accepted functionalism "whether a system has a mind, or how intelligent it is, is determined by what it can and cannot do."[66] What sort of input-output functions would a machine need to instantiate such that we would recognize it as fully conscious or intelligent? Alan Turing, one of the great geniuses of early computer science, proposed what has become known as the *Turing Test.* On this view, "a nonhuman system will be deemed intelligent if it acts so like an ordinary person in certain respects that other ordinary people can't tell … that it isn't one."[67] Turing agues that a system's linguistic or verbal behavior is the mark of intelligence: "A system is surely intelligent," he said, "if it can carry on an ordinary conversation like an ordinary person," setting aside obvious differences such as tone of voice, and the like.[68] Although we will encounter dissent in the following chapter, the Turing Test is taken by many philosophers and computer scientists as the criterion for intelligence (consciousness) on the part of any system, biological or mechanical. Moreover, if functionalism is the proper characterization of mind, we would expect that in principle computers capable of passing the Turning Test can be built, though we are certainly far from doing so as of now.

We have seen in this chapter a development of materialism from eliminativism to the various permutations of nonreductive materialism. We have found that any plausible version of materialism will most likely be a version of functionalism. Functionalism is a powerful theory. It seems to provide an account

66. John Haugeland, "What Is Mind Design?"in *Mind Design 2,* ed. J. Haugeland (Cambridge, Mass.: MIT Press, 1997), 3. Haugeland's essay is an excellent introduction to the philosophical and scientific issues involved in artificial intelligence debates.

67. Ibid. Turing's original essay, "Computer Machinery and Intelligence," appears in *Mind Design 2,* 29–56.

68. Haugeland, "What Is Mind Design?" 3.

of psychological states without even a hint of anything left to be explained by supposed nonphysical states, while at the same time allowing for mental causation without raising the problems of mind-body interaction. Functionalism also makes sense of the psycho-physical unity of human beings; since *human* minds are functional systems realized by human brains, it is not surprising to find that our consciousness is dependent on our neurophysiology, and it is clearly consistent with our being organisms and not pure spirits. Given that functionalism gives us a general theory of consciousness and intelligence that can be applied to any living or nonliving thing that operates in an appropriate way, it lacks the problems regarding the relationship between human and animal minds that plague dualism. In short, functionalism stands as a very formidable opponent to dualism. In the following chapter we will discuss how some philosophers have criticized this giant of contemporary ideas, but before doing so I hope the reader will stop to ponder the elegance and sheer genius of functionalism. It is, whatever its vices, *a very good idea*![69]

69. While completing this chapter, I have had the benefit of a great many comments and criticisms from an anonymous reviewer that have been particularly helpful. Whether the final version is satisfactory to him or her, I do not know. In any event, this chapter is much the better because of the care he or she took in commenting on the manuscript.

CHAPTER 5

PROBLEMS FOR MATERIALISM

In our last chapter we considered several philosophical approaches to psychological states that are at least consistent with materialism. On the one hand we considered eliminative materialism and logical behaviorism, but as we found those to be rather problematic, we will leave them aside in this chapter. On the other hand, we considered that various versions of the identity theory and functionalism enjoy a great deal more plausibility, and hereafter when I speak of materialism, I will have these sorts of positions in mind.[1] What these latter theories commonly claim is that when something has a sensation or thought, *that token* of a psychological state is a token of a physical state. According to token-identity theorists and functionalists, but not type-identity theorists, other tokens of that psychological type might have been realized by tokens of another neurophysiological state type (or possibly a nonphysical state type), and to explain this they use the notion of superve-

1. As we discussed in the last chapter, we should be careful when speaking about functionalism in the same breath as materialism, because functionalism is not a materialist theory per se (though it is consistent with materialism).

nience. Thus, whether he advances a classic type-identity theory, token-identity theory, or some version or other of functionalism, the materialist claims that an organism's (or a machine's) having a sensation or thought need not involve nonphysical properties or substances.

In this chapter, we will consider objections from philosophers who believe that these various materialist theories fail, because they all leave something essential out of the account of sensation, thought, and moral-intellectual agency; that is, there is some aspect of sensation, thought, or agency that is left unaccounted for by the materialist theories, and consequently we cannot say (at least for now) that psychological beings in the actual world are just physical beings. In what follows, we will first consider objections to materialism based on the qualitative nature of sensations. We will then consider objections based on the intentional nature of thought, and finally we will consider objections to materialism based on the nature of moral and intellectual agency. Note, however, that these objections do not necessarily amount to a rejection of naturalism, but only of materialism. As we shall see in the following chapter, there are versions of naturalism that concede that psychological states are neither strictly reducible to nor supervenient on physical states while claiming that psychological states can nevertheless be explained by underlying physical phenomena. Before we turn to nonmaterialist, naturalist positions, we first need to discuss whether there is any good reason to reject materialism.

Qualia Arguments—Bats, Cloistered Neuroscientists, and Zombies

Thomas Nagel points out that "if physicalism is to be defended, the phenomenological features [of psychological states] must be given a physical account."[2] The materialist must then

2. Thomas Nagel, "What Is It Like to Be a Bat?" reprinted in *The Philosophy of Mind: Classical and Contemporary Readings,* ed. David Chalmers (Oxford: Oxford University Press, 2002), 220. Originally printed in *Philosophical Review* 83, no. 4 (1974): 435–50; Nagel

construct an entirely physical story about sensations, but, "when we examine their subjective character it seems that such a result is impossible," because every sensation "is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view."[3] Nagel argues from the fact that the first-person perspective of sensations can be neither denied nor captured in any physical account. Philosophers call the qualitative and subjective aspects of sensations "qualia" (the plural of the Latin *quale* for "quality"), so objections to materialism like Nagel's are often called *qualia arguments.* He makes his case by appeal to the experiential life of the bat. Bats, as Nagel points out, are mammals with a very sophisticated perceptual apparatus, and there can be no doubt that they have experience, that is, there is some *way that it is like to be a bat,* in the sense that there is no way that it is like to be a stone. There is such a thing as bat awareness or bat feelings. Bats are primarily aware of the world through echolocation (a sort of biological sonar system); they locate prey and navigate by echoes rebounding off objects from the high frequency sounds the bats emit. Bats can *hear* (or something *like* hear) where otherwise silent objects are with such accuracy as to be able to capture insects in midflight. Although some blind people have learned a very rudimentary form of echolocation, what is it like to have such an experience is utterly inscrutable to anybody who has not had the experience. Whatever bat consciousness is like, it is utterly unavailable to the vast majority of human beings.[4]

Imagine now that Martha is the world's leading expert on bat neurology. In fact, imagine that Martha has a complete mastery

is using the term "physicalism" more or less in the same sense that we have been using the term "materialism."

3. Ibid.

4. Of course, we build sonar technologies that exploit the principles of echolocation, but that is not the same phenomenon as experiencing echolocation. The sonar technician doesn't echolocate his target directly, but sees a certain blip on a screen, which a machine has echolocated for him. The sonar technician is just as ignorant of the first-person experience of echolocation as any other human being.

of all that there is to know about the physical basis of echolocation in bats. Would this knowledge contribute in the least to Martha's understanding of *what it is like to be a bat*? Not at all; Martha would be no closer to understanding what it is like to have the sensation or quale of echolocation than somebody who completely lacks knowledge of bat neurology. The problem is that bat qualia are accessible only from the first-person perspective of the bat, and not from the third-person perspective of the neurologist. Since physical entities are accessible publicly, the inscrutability of bat qualia, even to the ideal expert on bat neurology, is relevant to our evaluation of materialism because "if the facts of experience—facts about what it is like for the experiencing organism—are accessible only from one point of view, then it is a mystery how the true character of experience could be revealed in the physical operation of the organism."[5] If bat qualia just are bat neurological states (or supervenient on bat neurological states), then we would expect to be able to derive our understanding of the former from the latter. Information about bat neurology, however, does not advance our understanding of bat qualia one *iota;* however much we advance in our understanding of the neurophysiological workings of bats, the qualitative aspect of bat sensations will remain utterly inscrutable to us non-echolocators. It is then difficult to see why we would ever conclude that bat qualia just are identical to or even supervenient on neurophysiological states of bats. For example, the nature of the statue of Dumbledore we discussed in the last chapter is perfectly transparent given the arrangement of its clay; that is, the arrangement doesn't seem "to leave anything out," but we can't say as much for the qualitative aspect of echolocation and the relevant states of the bat nervous system.

Frank Jackson, a self-avowed "qualia freak," makes a similar point using the example of a neuroscientist, Mary, "who is, for

5. Nagel, "What Is It Like to Be a Bat?" 222.

whatever reason, forced to investigate the world from a black and white room via a black and white television monitor."[6] Mary is an ideal expert, who has a complete knowledge of the physical basis of color perception; she lacks no information about what happens physically when we see a red object. Jackson believes that the possibility of somebody like Mary raises obdurate questions for the materialist:

> What will happen when Mary is released from her black and white room or is given a color television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. *Ergo* there is more to have than that, and Physicalism is false.[7]

Prior to leaving her black and white hermitage, Mary had no idea as to *what it is like to see red.* She would have new information presented to her following her release. One cannot doubt that there is such a thing as the experience of red, because, as Jackson's example shows, there is some difference in Mary's knowledge after she has encountered a red object. Once again, the fact that the qualitative aspects of perception are real experiences and utterly elude even complete physical knowledge shows that qualia and neurophysiological states are separate phenomena. We can see that there are such events as qualia to which we have direct access because we can contrast cases in which such episodes are present or absent, and this is what Jackson's (like Nagel's) example shows.

A way of making a similar point is a thought experiment using zombies. In this context, zombies are not the villainous monsters of *The Night of the Living Dead,* but precise physical duplicates of ourselves existing in possible worlds in which there are no qualia. My zombie, we'll call him *Zim Zadden,* is just like me

6. Frank Jackson, "Epiphenomenal Qualia," reprinted in *Mind and Cognition,* ed. William G. Lycan (Cambridge, Mass.: Blackwell Publishers, 1990), 471. Originally printed in *Philosophical Quarterly* 32 (1982): 127–36. Page references are to the Lycan volume. Jackson himself no longer completely endorses this line of thought, even though it continues to be very influential.

7. Ibid.

down to the arrangement of the quarks and electrons, but Zim Zadden lacks any qualia at all; for example, he does not see red, taste habanera sauce, feel a runner's high, even though we are physically the same top to bottom. We might say that Zim Zadden and I would both realize the same causal-functions, but the former lacks any of the associated qualitative experiences had by the latter. Philosophers such as Nagel and David Chalmers believe that the existence of Zim Zadden is certainly possible; that is, there is a possible world in which Zim Zadden exists. If we admit that, however, we must likewise admit that there is a difference between Jim Madden and Zim Zadden, that is, the seeing red, tasting habanera sauce, feeling a runners' high, and so on. If the proponent of zombies is correct, then these psychological states do not supervene on physical states, because Zim Zadden would be physically indiscernible from me and yet psychologically different.[8]

Nagel summarizes the point of these arguments when he claims that

> the idea of how a mental and a physical term might refer to the same thing is lacking, and the usual analogies with theoretical identification in other fields fail to supply it. They fail because if we construe the reference of mental terms to physical events on the usual model, we either get a reappearance of separate subjective events as the effects through which mental reference to physical events is secured, or else we get a false account of how mental terms refer.[9]

8. An early version of a "zombie" argument can be found in Thomas Nagel's "Armstrong on Mind," in *Readings in Philosophy of Psychology,* ed. Ned Block (Cambridge, Mass.: Harvard University Press, 1980), 200–206. For a more recent version of the zombie argument see David Chalmers, "Consciousness and Its Place in Nature," in *Readings in the Philosophy of Mind: Classical and Contemporary Readings,* ed. David Chalmers, 242–72 (New York: Oxford University Press, 2002), 247–72; and *The Conscious Mind.* Edward Feser provides a very helpful discussion of the controversy surrounding the possibility of zombies in *Philosophy of Mind: A Beginner's Guide* (Oxford: Oneworld Publications, 2005), 93–96. See Gutting, *What Philosophers Know,* 125–38, and Scott Soames, *Reference and Description,* (Princeton, N.J.: Princeton University Press, 2005), 194–266, for technical criticism of Chalmers's use of zombies.

9. Nagel, "What Is It Like to Be a Bat," 224. Jackson is careful to distance his own argument from Nagel's for reasons ancillary to our concerns. See Jackson, "Epiphenomenal Qualia," 472–73.

Nagel's point is that we can use strict reduction to make sense of standard scientific reductions, for example, "Water is H_2O," because we can understand how water could turn out to be H_2O, or in some cases we are willing to replace one theoretically useless entity (e.g., phlogiston) for a theoretical fruitful one (e.g., oxygen). When we reduce water to H_2O or replace phlogiston with oxygen, we don't seem to leave anything out; in both cases our new description is seemingly complete. As I point out above, we encounter a similar problem for supposedly nonreductive materialism; in straightforward cases of supervenience there isn't an explanatory gap, whereas we find just such a breach in supposed cases of psycho-physical supervenience. The problem for materialists is that when we attempt either to reduce psychological states to physical states or to argue that they are supervenient on physical states, we seem inevitably to leave something important out of our final story about sensation, or at least that is what Jackson's, Nagel's, and Chalmers's thought experiments purport to show.

Materialist Replies to the Qualia Arguments

The arguments based on qualia, especially Jackson's, often called knowledge arguments, are among the most powerful arguments against materialism, and they have spawned a vast literature.[10] We will not be able to address all, or even most, of the criticisms of the knowledge arguments (let alone its defense by various proponents). We will have to settle for discussing a few of the most important objections. The most common, even standard, objection to the so-called knowledge argument is to accuse its proponent of a gross logical fallacy. Andrew Melnyk offers a

10. In this chapter I am presenting what I call the qualia arguments primarily as objections to materialism, but they are frequently purported to be significant arguments in favor of dualism, property dualism in particular. The reader may do well to consider these arguments alongside the arguments presented in chapter 2 when assessing the complete case for dualism.

clear presentation of this reply on the part of the identity theory (using the example of pain as a quale):

> I think the Knowledge Argument fails, because it relies on a general principle that's false. According to the Knowledge Argument, since Mary knows all the science of pain, but then learns something new (i.e., that pain is like that), it must be that the subject matter of the new knowledge (i.e., pain's being like that) is a feature of reality that the science of pain didn't refer to or talk about at all. The general principle underlying this reasoning seems to be this: if someone knows a certain body of knowledge, but then learns something new, it must be that the subject matter of the new knowledge is a feature of reality that the old body of knowledge didn't refer to or talk about at all. However, this general principle seems false; it holds true of very many but not all situations.[11]

Melnyk's point is really a reiteration of the original point that motivates the identity theory, that is, there is nothing philosophically suspect about two conceptually distinct descriptions referring to one and the same object. He grants that Mary does not know anything about what it is like to see red prior to leaving her black and white study, so she certainly does learn something new when she finally sees red. That, argues Melnyk, does not necessarily show that seeing red and its corresponding neurophysiological state (which Mary understood before seeing red) are really distinct. Counterexamples that make Melnyk's point are easy to generate. Melnyk asks us to consider a scenario wherein he (Melnyk) suffers from amnesia and subsequently forgets his own name. He reads in a newspaper that somebody named "Andrew Melnyk" is to receive corporal punishment tomorrow, but this is not in the least distressing because he believes that he and this Melnyk character are distinct individuals. He later discovers that he is indeed Andrew Melnyk. After learning that he is to be flogged, Melnyk has new and significant knowledge, but

11. Andrew Melnyk, "Naturalism, Free Choices, and Conscious Experiences," accessed at http://www.infidels.org/library/modern/andrew_melnyk/against-dualism.html on 7/23/09. Stewart Goetz and Charles Taliaferro reply to Melnyk's attack on the knowledge argument in "Reply to Melnyk's Objections," accessed at http://www.infidels.org/library/modern/stewart_goetz/against-melnyk.html on 7/23/09, and in their book *Naturalism,* 48–50.

this doesn't mean that there is some separate entity to which this new knowledge refers. It was just Melnyk all along. Sometimes we discover identities, even counterintuitive identities. In short, just because I can gain new information about an object, doesn't imply that this object is not one but two.

Melnyk is perfectly correct on this point; that is, two conceptually distinct descriptions can in fact refer to a single, self-identical object. Thus, the fact that somebody gains the ability to give a novel type of description does not mean that she is now acquainted with a novel type of object. The problem for Melnyk, and the standard reply to the knowledge argument in general, is that Nagel's and Jackson's arguments need not be construed so as to fall into such an obvious fallacy.[12] The point of the so-called knowledge argument is not that a descriptive (conceptual) difference implies different objects of knowledge, but that we can account for the difference in Mary's newly acquired knowledge of *what it is like to see red* only in terms of a qualitative experience she has for the first time after leaving the room. Likewise, we can account for our lack of knowledge of what it is like to be a bat only because of our lack of the experience of echolocation. The key to the Nagel-Jackson argument "is a point about what real features exist in the world and not, except derivatively, about how we know about those features."[13] Searle makes this point in more detail when he argues:

> The point of the argument is that there exist real phenomena that are necessarily left out of the scope of their knowledge, as long as their knowledge is only of objective, third-person, physical facts. The real phenomena are

12. Goetz and Taliaferro make a different set of replies to Melnyk: (a) They argue that there is a disanalogy between Jackson's Mary example and Melnyk's amnesia example; i.e., in the latter Melnyk had experience of himself and just didn't know it, whereas in the former Mary had no experience of red; (b) they offer a version of the difference argument appealing to the simplicity of mental states and the complexity of brain states. I find (a) compelling and will endorse a version of this position developed by Searle, though I think that at this point in *our* discussion (b) begs the question. See Goetz and Taliaferro, *Naturalism,* 49.

13. Searle, *The Rediscovery of the Mind,* 117.

color experiences and the bat's feelings, respectively; and these are subjective, first person, conscious phenomena. The problem in Mary's case is not just that she lacks information about some other phenomenon; rather there is a certain type of experience that she has not yet had. And that experience, a first-person subjective phenomenon, cannot be identical with the third-person, objective neuronal and functional correlates. The point about epistemology, the information, is just a way of getting at the underlying ontological difference.[14]

The so-called knowledge argument doesn't trade on an inference from the fact that psychological states can be known by two different descriptions to the conclusion that they must involve both nonphysical and physical properties. Rather, the point is that we can know that certain descriptions apply to a particular psychological state only by having the requisite experience. We need to posit a *qualitative experience* as a real feature of the world to account for the difference. If we say that experiences are not real, that they are just brain processes, or that they supervene on those processes, we cannot account for the difference in Mary's knowledge or our lack of knowledge of bat qualia. Descriptive differences do not tell us much about what exists, but if we know that a certain description is true of some object, and the only account to be given of that knowledge is in terms of some experience, then we must grant the reality of the *experience.* Since there is seemingly no room for first-person, qualitative experiences in physical reality, materialism is then clearly incomplete as an account of our sensations. It seems that if she admits the reality of such states as *what it is like to see red* and *what it is like to be a bat,* then the materialist must suffer the charge of at least property dualism.

Another avenue of response to the knowledge argument is to argue, along with Laurence Nemirow, that "knowing what it's like may be identified with knowing how to imagine."[15] In other

14. Searle, *Mind,* 67–68.

15. Laurence Nemirow, "Physicalism and the Cognitive Role of Acquaintance," in *Mind and Cognition,* ed. William Lycan (Cambridge, Mass.: Blackwell Publishing, 1990), 493.

words, when Mary leaves her black and white study, she does not gain a first-person, irreducible, qualitative state, but she does gain something, that is, an *ability* to imagine, recognize, and describe red that she lacked before. Nemirow presents the advantage of the ability hypothesis when he claims:

> The more seriously we take this ability question, the easier it becomes to resist the knowledge argument. The latter assumes that science cannot convey what it's like to see red. The premise is uncontentious, for science does not seek to instill imaginative abilities. But the knowledge argument concludes that physical science cannot describe certain information about seeing red. The inference is invalid because it presumes that to know what it's like is propositional knowledge rather than an ability.[16]

Nemirow argues that to know a quale is not to know a fact, as when I know where my pocketknife is located. Rather, understanding a quale is a sort of implicit know-how, a power or ability to make a certain performance. Know-how is often something that defies third-person, objective description; for example, the know-how necessary for riding a bicycle is extremely difficult, if not impossible, to convey objectively, but it doesn't seem to pose any problem for materialism. To say that Mary knows *what it is like to see red* is really just to say that she can do something, namely imagine or describe the color red, and to say that Martha lacks knowledge of *what it is like to be a bat* or that Zim Zadden wouldn't know *what it is like to taste habanera sauce* is to say that they lack the ability to imagine echolocation or the combination of something extremely spicy yet fruity. If Nemirow is correct then, as David Lewis puts it, "phenomenal information [qualia] is an illusion," though we can explain this illusion in terms of an easy confusion between *knowing how* and *knowing that*.[17] We can admit that Mary gains something and zombies lack something, but what is lost or gained is not an experience or irreducible quale, but an ability to imagine or describe.

16. Ibid.
17. David Lewis, "What Experience Teaches," in Lycan (ed.), *Mind and Cognition,* 517.

When considering this defense of materialism, there is no need for the proponent of the knowledge argument to deny, for example, that Mary gains new abilities to imagine and describe, but it must also be pointed out that she has gained the ability to imagine and describe her new knowledge. As Kim puts it, upon release "Mary certainly gains these abilities, but that does not preclude her also gaining propositional knowledge, knowledge of facts about how things look to her and other people."[18] That is, no doubt Mary gains a new ability to imagine and describe, but what she can now imagine and describe, say the Qualia Freaks, is what it feels like to see red, something of which she was previously ignorant. The ability hypothesis does not address the issue of what exactly Mary is imagining or describing, that is, her first-person experience of red. The materialist might attempt to account for imagination or description in terms of behavioral dispositions, various psycho-physical identities (type or token), or causal-functional roles, but this only pushes the problem back one remove; in each case the proponent of the knowledge argument would just as likely ask whether these accounts actually capture what Mary imagines or describes, that is, her awareness of what it is like to see red. Thus, the ability hypothesis doesn't pose a great threat to the knowledge argument.[19]

Daniel Dennett argues that we really have no choice but to eliminate qualia from our account of sensations. Dennett is quite willing to embrace the apparent ironies inherent in denying qualia: "At first blush it would be hard to imagine a more quixotic quest than trying to convince people that there are no such properties as qualia.... But I am not kidding."[20] As we have seen

18. Jaegwon Kim, *The Philosophy of Mind*, 3rd. ed. (Boulder, Colo.: Westview Publishing, 2011), 324.

19. This is, admittedly, a quick and simplified critique of the ability hypothesis, which is a very influential proposal. For further discussion and citations of more of the literature both pro and con regarding the ability hypothesis, see Kim, *The Philosophy of Mind,* 323–26.

20. Daniel C. Dennett, "Quining Qualia," in *Mind and Cognition,* ed. William Lycan (Cambridge, Mass.: Blackwell, 1990), 519.

above, the qualia arguments proceed by way of examples to motivate our intuitions regarding our ordinary experiences, or lack of experiences. Likewise, Dennett replies to the qualia arguments by "intuition pumps" or thought experiments of his own that in turn show that the very notion of a quale "is so thoroughly confused" that it would be "far better … to declare that there simply are no qualia at all."[21]

Dennett uses fifteen different examples in his paper to show what he believes to be the inherent confusions regarding qualia. I will not address all of these, but only the one that seems the most representative of Dennett's general case. He asks us to consider the case of Chase and Sanborn, two professional coffee tasters employed by Maxwell House for the last six years. One day Chase confides in Sanborn that he no longer likes the taste of Maxwell House coffee. Chase believes that the taste has remained the same, though he simply doesn't like it anymore. Sanborn uses this occasion to make a similar confession: he no longer likes Maxwell House, though he believes that the taste of the coffee has changed. That is, whereas Chase believes the taste of Maxwell House has remained constant while his attitude toward it has changed, Sanborn believes that something has gone wrong with his tastes buds or the Maxwell House recipe such that the coffee tastes different to him.[22]

Dennett argues that this logically possible scenario raises serious problems for qualia. According to Chase, his quale, that is, the taste of Maxwell House, remains constant, but how can he be sure of that? It seems that he may very well be in the same predicament as Sanborn, without any difference in appearance to him at all. How does one know that Maxwell House tastes today the same way as it did six years ago? We can make the same point regarding Sanborn; for all he knows, contrary to his own intuitions, he might be in the very same predicament as Chase, that

21. Ibid., 520.
22. Ibid., 527.

is, the taste of Maxwell House has remained the same, though his attitude toward it has changed. The point is that neither Sanborn nor Chase can be certain as to what his qualia are; they don't know whether they are having the same *tasting states* as they did six years ago. No introspective analysis of how things feel can solve this problem for Chase and Sanborn, though a neurological study of both men and a chemical analysis of the coffee could go a long way toward resolving the dispute. Dennett believes that this poses a problem for qualia because "the idea that one should consult an outside expert, and perform elaborate behavioral tests on oneself in order to confirm what qualia one had, surely takes us too far away from our original idea of qualia as properties with which we have a particularly intimate acquaintance."[23] Since the discrepancy between Chase and Sanborn cannot be solved by any reflection on their internal, private experiences, but it could be addressed by an objective, third-person investigation, we have reason to doubt that supposed qualia do any work at all in explaining behaviors and attitudes.

Dennett argues that, since qualia do no apparent explanatory work, there "is no reason to be cowed into supposing that [sensations] have some intrinsic properties behind, or in addition to, their various dispositional, reaction-provoking responses" and that "qualia are no more essential to the professional vocabulary of the phenomenologist (or professional coffee taster) than to the vocabulary of the physiologist."[24] That is, the notion of qualia over and above what can be given a functional characterization in terms of a causal role linking stimuli with responses is useless and confused. Sensations are not entities that come to us with intrinsic, qualitative properties (qualia), but are topically neutral episodes we posit as part of an explanatory theory of behavior. If a functionalist account of sensations can more parsimoniously explain such behavior, there is no reason to concede obscure epi-

23. Ibid., 533.
24. Ibid., 534–35.

sodes like qualia. We do best then to dispense with any notion of sensation that goes beyond what can be functionally characterized in our theory of mind. If we mean by qualia properties that cannot be functionally characterized, then "contrary to what seems obvious at first blush, there simply are no qualia at all."[25]

Early in his presentation of this attempted elimination of qualia, Dennett entertains a likely objection:

> There is a strong temptation, I have found, to respond to my claims in this paper more or less as follows: "But after all is said and done, there is still something I know in a special way: I know how it is with me right now." But if absolutely nothing follows from this presumed knowledge … what is the point of asserting that one has it? Perhaps people just want to reaffirm their sense of proprietorship over their own conscious states.[26]

The point of the objection is that even if neither Chase nor Sanborn can confirm definitively whether his qualia are the same regarding Maxwell House coffee, there is still some way that it tastes to them now. They may be confused as to what their qualia are, in comparison to some past experience, but they nevertheless have qualia. In short, *having a confused experience is nevertheless having an experience.* Thus, showing that our notions of qualia might be confused in cases like Chase and Sanborn does not dispel the apparent reality of qualia.

Galen Strawson makes this point well when he says that, though Dennett might succeed in showing that qualitative consciousness is confused or noncognitive, he fails "to put into question its existence or reality. Whatever process by which the seeming arises, the end result of the process is … at least this: that it seems as if one is having phenomenally rich experience of Beethoven's eighth quartet or an Indian wedding. And if there is this seeming, then, once again, there just is phenomenology or experience."[27] Dennett's reply is to wonder exactly what qualia ex-

25. Ibid., 544.
26. Ibid., 528.
27. Strawson, *Real Materialism and Other Essays,* 55. See also Taliaferro and Even's discussion of Dennett on these issues in *The Image in Mind,* 89–91

plain, given that they do not seem to explain anything. Qualia do not make any difference for our attitudes and behaviors, so there is no reason to posit them in our account of sensations. This argument, however, "presupposes that only behavior needs explaining. The opponent will hold that qualia are an explanation in their own right."[28] In other words, Dennett assumes that there is nothing more to explain than our behavioral responses to stimuli, and not our intrinsic conscious experiences. The proponents of the various qualia arguments claim to show that qualia are the basic, irreducible data of experience. They posit them not to explain our behaviors, but to articulate an undeniable fact of consciousness: that is, there are ways that things feel to us. Dennett assumes "that the only 'seemings' that need explaining are dispositions to react and report," but that is not to defend functionalism, but merely to assert it outright.[29] Dennett's argument then gives us little reason to doubt the various arguments from qualia against materialism.

Arguments from Thought—Beliefs, Intentionality, and the Man in the Chinese Room

Let's leave the issue of sensations aside for the moment and focus our attention on arguments against materialism based on our other major class of psychological events, that is, thoughts. In this section we will mainly consider objections to functionalist accounts of thought. As we have discussed earlier, materialism and functionalism must be distinguished (one could be a materialist without being a functionalist and vice versa), so a defeat of functionalism isn't necessarily a defeat of materialism. Nevertheless, we also found that it is likely that the materialist's best bet for giving an account of psychological states involves adopting functionalism. Thus, if functionalism were utterly inadequate as an account of thoughts, we would have a serious strike against the plausibility of materialism in its strongest contemporary form

28. Chalmers, "Consciousness and its Place in Nature," 252.
29. Ibid.

(even if this would not amount to a demonstrative refutation of materialism per se). We will then proceed in this section assuming that a good reason to believe that functionalism cannot in principle account for thought is likewise a good reason to doubt materialism.

You will remember from chapter 1 that one of the hallmarks of thought is *intentionality.* That is, our thoughts *refer to* or are *about* certain objects or states of affairs inasmuch as they consist of true or false assertions about objects or states of affairs. For example, *Patrick thinks that Martha is beautiful* means that Patrick makes a claim about an object, Martha; his belief refers to or is about Martha. Another way we might think of Patrick's thought is as a representation; by thinking that Martha is beautiful, Patrick represents her as having the property *beautiful.* Other philosophers prefer to think of intentionality in terms of *propositional content* or the information included in a thought, but information is always *information about something,* an intentional concept. It is extremely difficult to have anything very informative to say regarding what we mean by such words as "about," "refers to," "represents," "content," or any other word we might use to characterize intentionality, but we certainly must accept that our thoughts have intentionality. If our thoughts did not refer to or weren't about something or didn't convey information, then it seems that they could not be either true or false. In other words, it is in virtue of their intentionality that our thoughts can get the world either right or wrong. Thus, any account of thought, materialist or otherwise, must include intentionality.

We should also note that intentionality is a very strange phenomenon. The mind is able to have thoughts about objects that are not only present but also absent: for example, I can have thoughts about Phoenix, Arizona, while I'm sitting in Kansas. We can even have thoughts about nonexistent objects. For example, William can think about fictional characters such as hobbits and hippogriffs, even though no such thing has ever existed. More-

over, minds have an ability to bestow intentionality on ordinary inanimate physical objects. For example, suppose that the following set of symbols appeared on a blackboard:

(1) Napoleon won at Austerlitz.

This sentence would have intentionality, but only *derivatively*. That is, the streaks of chalk particles that compose (1) as written on a blackboard have intentionality (they are about Napoleon and the city of Austerlitz) only because the history professor and students together ascribe a shared understanding or meaning to these particular symbols. If by some highly improbable series of events, a mass of chalk dust were blown by the wind into the shape of the symbols in (1), and no intrinsically intentional agent existed to interpret the formation of chalk dust as meaning *Napoleon won at Austerlitz* (or any other meaning), this particular set of marks would not have intentionality. Utterances and inscriptions have intentionality, but it is derived from the intentionality of the speakers of the language in which they are made. A series of rabbit tracks in the snow, the wind blowing in the trees, a sunrise, and the like might have intentionality, if some group of people, for example, ascribe a certain cultural or religious significance to such an occurrence, but once again this intentionality is derived or extrinsic; that is, it is given to the marks or symbols from the outside.[30] The intentionality of our thoughts, however, is not extrinsic but intrinsic. Our thoughts are not intentional because something external to us interprets them; they just are intrinsically intentional. In any event, it can't be the case that all intentionality is extrinsic. We could not go on borrowing intentionality to infinity, so at

30. Rabbit tracks, the blowing of the wind in the trees, and the sunrise are examples of what are often called natural signs; e.g., rabbit tracks in the snow are a natural sign that a rabbit has recently passed this way. I agree that natural signs might be intrinsically so. That the blowing of the wind in the trees may signify the onset of autumn is not derivative from anything else. However, the relationship between a natural sign and the event or object it signifies is causal, not intentional. Natural signs have intentionality or meaning only if they are interpreted by some intrinsically conscious agent. See Peter King, "Medieval Intentionality and Pseudo-Intentionality," *Question* 10 (2010): 25–44, for a discussion of certain medieval conceptions of nonderivative intentionality in nature.

some point there must be something that is intentional in itself, if there is to be intentionality at all.

We will discuss typical functionalist strategies for dealing with the intentionality of thought a bit later, but now I want consider an argument against functionalism that takes intentionality as its point of departure. John Searle has defended a thought experiment he calls "The Chinese Room" in various publications over the last two decades as part of an argument against functionalism.[31] You will remember that one of the key theses of functionalism is that all psychological states can be analyzed in terms of functional roles that link stimuli from an environment with behavioral responses. That is, for the functionalist, whatever it is that links an environmental input with a behavioral response is a token of a psychological state of a certain kind. For example, William's believing that water is wet is the neurophysiological state that links an input (somebody asking William, "Is water wet?") and a behavioral response (William's reply, "Yes"). In human beings this link is provided by a neurophysiological state of a certain kind, though in other organisms a different type of neurophysiological state may play this role. In fact, there is no reason to limit the physical instantiation of a functional role to "wet circuits" like neurons, so the role need not be played by a neurophysiological state at all. If a computer, according to a very strong functionalist story, has internal states that link inputs with appropriate outputs, we should characterize those internal states as psychological states. They serve the same function as certain psychological states, and,

31. Searle's first presentation of the Chinese Room is in "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 1 (1980): 417–24. He has developed and defended this position in *Minds, Brains, and Science* (Cambridge, Mass.: Harvard University Press, 1984), 28–41; *The Rediscovery of the Mind,* 44–45; and *Mind: A Brief Introduction,* 62–64, 69–71. There are several very powerful arguments against materialism that take intentionality as the point of departure that can be found in the classical tradition of philosophy. I have not reproduced any such arguments in this chapter and instead have relied upon Searle's argument alone. This decision is made not because I don't think such arguments are compelling, but because they can be equally effective against emergentist versions of naturalism(including Searle's own view), which I will discuss in the next chapter.

for the functionalist, functional equivalence amounts to psychological equivalence. Thus, functionalism seems to imply some brand of what Searle calls strong artificial intelligence; if a machine functionally equivalent to human beings could be constructed, then such a machine would have thoughts in the same sense that we do.

John Searle's famous Chinese Room example is designed to show that functional equivalence is not sufficient for psychological equivalence. Searle explains this example:

> Imagine that someone who understands no Chinese is locked in a room with a lot of Chinese symbols and a computer program for answering questions in Chinese. The input to the system consists in Chinese symbols in the form of questions; the output of the system consists in Chinese symbols in the form of answers to the questions. We might suppose that the program is so good that the answers to the questions are indistinguishable from those of a native Chinese speaker. But all the same, neither the person inside nor any other part of the system literally understands Chinese; and because the programmed computer has nothing this system does not have, the programmed computer, qua computer, does not understand Chinese either. Because the program is purely formal or syntactical and because minds have mental or semantic contents, any attempt to produce a mind purely with computer programs leaves out the essential features of the mind.[32]

The man in the Chinese Room would function just like a digital computer; he would have a preprogrammed system linking inputs with appropriate outputs. Moreover, if we assume that his manual is sufficiently detailed, we would expect the man, or maybe the entire Chinese Room system, to be functionally equivalent to any native, human speaker of Chinese; that is, the man could answer questions to a degree of accuracy indistinguishable from a native Chinese speaker, so that he could pass the appropriate Turing Test. Searle's point is, however, that the man doesn't understand a word of Chinese! The man in the Chinese Room doesn't know what the Chinese outputs are about, so he has *at most* extrinsic intentionality. That is, although he can give

32. Searle, *The Rediscovery of the Mind,* 45.

syntactically correct replies to Chinese questions, his Chinese utterances lack semantics; that is, they have no intentionality, unless they are interpreted by the Chinese speakers outside the room. Functionalists and proponents of strong artificial intelligence argue that having beliefs just is a matter of instantiating the right sorts of functional roles, but that means that we are all like the man in the Chinese Room, even when speaking our native tongues. The man (or the entire system of the Chinese Room) lacks intrinsic intentionality, though it is functionally equivalent to intrinsically intentional systems, such as a human speaker of Chinese. Thus, if functionalism is correct, all intentionality is extrinsic, but that result is absurd.

Furthermore, we would not say that the man, nor the entire Chinese Room system, has thoughts corresponding to his Chinese outputs. His outputs lack intentionality, and thoughts must have intentionality. It then seems that functionalism, to borrow Searle's phrase, "leaves out" one of the essential features of thought, namely intentionality. Of course, systems like computers might have extrinsic intentionality; that is, their symbolic outputs have meaning because we interpret them as such, but our intentionality must be intrinsic. Thus, a functionalist characterization like what we have discussed so far does not capture all of the essential features of thought. The point then is that, given the Chinese Room example, it seems that even an exhaustive functional analysis would leave out intentionality, and therefore thoughts. We have good reason then to conclude that a functionalist theory of mind fails.

Can functionalism be refined so as to account for intentionality, the missing ingredient for thought? Certainly, functionalists do not take these objections lying down, and some prominent strategies have arisen for providing an entirely materialist account of intentionality. Maybe the most common strategy is to account for intentionality in terms of the causal relationship between an external object (or event) and the functional state in-

ternal to the organism having the thought. That is, a state of an organism is about or refers to an object inasmuch as it is caused by the object and it disposes the organism to react to the object, even if that disposition is not actually exercised.[33] According to a causal theory of intentionality, Patrick's thought (which turns out to be a certain functionally characterized neurophysiological state) is about Martha, because Martha plays a causal role in bringing about this particular state in Patrick and it disposes Patrick to react to Martha. On this view, intentionality is no more mysterious or immaterial than any relation of physical causation; to say that a certain disposition is about a particular object is simply to attribute to that object a role in causing it.

It is fairly easy to see several problems with reducing intentionality to a causal relationship. It is certainly not a necessary condition for having a thought about an object that such a thought is caused by that object. It is possible to have a thought about a certain state of affairs that is not caused by such a state of affairs; for example, we can have thoughts about nonactual or even impossible states of affairs, which certainly cannot be caused by their objects. Maybe such thoughts must be derived from thoughts about objects we have directly experienced, but that doesn't imply that thoughts about nonactual objects are about the object of those experiences; for example, William's thought about a unicorn isn't about the ponies and horned animals he has experienced. Moreover, similar examples show that a causal relation between an object and a thought is insufficient to determine intentionality; for example, a causal relation with William might lead me to have a belief about Patrick in a case of mistaken identity.

33. This is a greatly simplified and generic version of a cluster of highly complex theories among which there are crucial differences. For a useful introduction to causal theories of intentionality, see Armstrong, *The Mind-Body Problem,* 137–44. See also Tim Crane, *The Mechanical Theory of Mind* (London: Penguine Books, 1995); Fred Dretske, *Naturalizing the Mind* (Cambridge, Mass.: MIT Press, 1995); and Jerry Fodor, *A Theory of Content and Other Essays* (Cambridge, Mass.: MIT Press, 1990).

A further problem for causal theories is revealed if we suppose that Brendan is confronted by a blue, cylindrical cup that enters into a causal relation with his sensory faculties causing him to think at a time, *t1*:

(2) The cup is blue.

Further, suppose that at *t2* Brendan thinks:

(3) The cup is cylindrical.

Propositions (2) and (3) express different thoughts; they have different *content,* because they are about different properties of the cup. There is no reason to suppose that the causal relationship between Brendan and the cup is different at *t1*and *t2,* nor is there any reason to suppose that Brendan's disposition to react to these aspects of the cup is different. That is, the fact that Brendan thinks (2) at *t1* and (3) at *t2* in no way implies that the cup is exerting any different causal influence on his sensory apparatus or his behavioral dispositions. The cup's position, surface texture, Brendan's perspective, and other factors all can remain constant while Brendan moves from thinking (2) to thinking (3). The point is that Brendan can have two different thoughts involving the very same causal relation with a physical object. The propositional content, meaning, or intentionality of a thought is therefore not accounted for by the causal relationship with the object of the thought.[34]

Even if we set that problem aside, it is still very difficult to understand how brute, physical causal relations could result in intentionality. When two billiard balls collide, no intentionality results; the eight ball's motion is in no straightforward sense *about* or *means* the collision it just had with the cue ball.[35] Physical

34. Searle and John Haldane make similar objections to materialism in *Mind: A Brief Introduction,* 65–66, and "A Return to Form in the Philosophy of Mind," in *Form and Matter: Themes in Contemporary Metaphysics,* ed. David Oderberg, 40–64 (Oxford: Blackwell, 1999), respectively.

35. This is not to say that the motion of the eight ball is not a natural sign of the collision.

causation alone is not an intentionality-endowing relationship, especially when there is no intentionality in the cause. According to the causal-functionalist story about intentionality, nonintentional physical causes somehow result in neurophysiological (or computational) states that have truth/falsity, meaning, reason, and so on. The problem is that no story about physical causation is sufficient to explain such intentional states. For example, one might be subject to all of the causal conditions sufficient to bring about the neurophysiological event associated with the belief *Water is wet* and even have the appropriate, accompanying dispositions, but "still not believe that water is wet. This is just an extension of the Chinese Room argument, but the moral it points to is general: You cannot reduce intentional content (or pains or 'qualia') to something else, because if you could they would be something else, and they are not something else."[36] Of course, a functionalist might simply assert that meaning, intentionality, truth, and the like just are certain causal-functional roles, but that is to beg the question. The Chinese Room thought experiment shows that a functional analysis of thought leaves something out, and this very same explanatory gap reappears in the causal story about intentionality. To say that intentionality just is a special case of brute, physical causation, is to leave intentionality out of the account.

Not all materialists attempt a straightforward reduction of intentionality to causality. One popular alternative strategy is to appeal to the success of Darwinian explanations in biology. For example, Ruth Millikan argues that a functional state refers to or is about a certain object just in case it has evolved in order to help the organism to obtain or react to such an object. That is, an internal state of an organism is about or refers to an object because it has evolved in order to help the organism to react to similar phenomena. Intentionality is not here identified with the causal

36. Searle, *The Rediscovery of the Mind,* 51.

relation between the dispositional state of the organism and an external object alone (though the causal link may still have some necessary role in determining intentionality), but also includes the purpose or function that such a state plays in the economy of survival for members of the organism's species. A certain functional state counts as a reaction to an object inasmuch as it has enhanced the organism's adaptive success in dealing with such an object. The intentionality of our thoughts is then a product of the roles they (or similar thoughts) have played in the prior evolutionary history of the human race. For example, a particular neurophysiological state in Martha has intentionality with respect to fire because it has evolved in the human species to serve the purpose of avoiding burn injuries.[37]

A likely objection is the fact that we all have thoughts for which there seems to be no readily apparent, or even broadly plausible, evolutionary advantage.[38] For example, I believe that *The Hobbit* is the best of Tolkein's novels, but it strains credulity to claim that my ability to have such a thought is somehow accounted for in terms of my physical survival. The situation seems even worse when we note that our neurophysiological equipment evolved hundreds of thousands of years ago to suit the needs of our early ancestors on the sub-Saharan plains of Africa. What possible role could a capacity for beliefs about the superiority of *The Hobbit* play in the survival of such organisms? That is, it is indeed dubious that the ability to have thoughts about fictional characters or logical abstractions would offer much in the way of direct survival value to our prehistoric ancestors who were primarily concerned with avoiding prehistoric predators

37. See Millikan, *Language, Thought and Other Biological Categories;* "Biosemantics," in *The Philosophy of Mind: Classical and Contemporary Readings,* ed. David Chalmers, 500–508 (New York: Oxford University Press, 2002); and Fodor, *A Theory of Content and Other Essays.* Feser provides a very helpful discussion of positions along the lines of Millikan's in *Philosophy of Mind,* 184–89.

38. Searle and Alvin Plantinga both make this point. See Searle, *The Rediscovery of the Mind,* 50; Plantinga and Tooley, *Knowledge of God,* 65–66. Feser likewise raises this point in the discussion I mention in the previous note.

such as saber-toothed tigers while garnering the romantic favor of potential mates. A proponent of the evolutionary account of intentionality might point out that the ability to have such apparently useless thoughts (at least for our primordial survival needs) could be a side effect of other modes of thinking that did indeed have a great deal of survival value. That is, presumably an ability to refer to or think about (and thereby "get right") one's environment (and the benefits and dangers therein) would be of some help to our ancient ancestors, and these abilities might further allow us to have thoughts about the relative merits of *The Hobbit*.[39] Nevertheless, one might worry that, given the evolutionary account of intentionality, my thought supposedly about the superiority of *The Hobbit* is indeed actually about the phenomena that the relevant ability was evolved to address, which seems rather odd; my thought would not really be about the novel, but about phenomena deep in our evolutionary past. Thus, even given the possibility of felicitous side effects in the development of intentional abilities, our capacity to think about objects far outstrips what can be accounted for in evolutionary terms. The history of our species may well tell us a great deal about *how* and *why* we developed some of our intentional capacities, but it sheds no light on *what* intentionality is essentially.

A further objection to evolutionary accounts of intentionality has been raised by Hilary Putnam. Suppose that a certain organism has an internal state that has played an adaptive role in the evolutionary history of its species, for example, a dog's neurophysiological state associated with meat. Putnam argues that the adaptive role of the dog's neurophysiological state with respect to meat is insufficient to make it such that the dog has a *thought about* meat.

39. Thanks to an anonymous reviewer for raising the plausibility of this sort of proposal. Plantinga argues against this sort of position in *Where the Conflict Really Lies*.

I may decide that one of my thoughts was successful in enabling me to maximize my well-being, but was not in fact true. I was deceived, but the deception was fortunate. No such cognitive performance is possible in the case of the dog. For a dog, the very distinction between having a true belief and having a successful belief simply does not make sense; and that means that the notion of dog's thought as being true or false, and of its proto-concepts as referring or not referring to something, simply do not make sense.[40]

Leaving aside the issue of whether nonhuman animals have bona fide thoughts, Putnam's point is that *having an adaptive role with respect to x* is not the same as *being about x* or *referring to x.* That is, an internal state of an organism could have an adaptive role with respect to some object, and fail to represent that object. Putnam's dog, for instance, has a neurophysiological state that is associated with meat, and certainly that state endows the dog with an adaptive advantage (it's definitely good for the survival and reproduction of dogs that they eat meat), but certainly even if that state did not represent meat or did not refer to meat, it would be adaptive so long as it led to survival and reproductive success. Indeed, as Putnam points out above, an organism might even have an internal state that plays an adaptive role with respect to some object, while at the same time failing to represent or even misrepresenting it. This is not to deny that we might have intentional states like thoughts and beliefs because they play an adaptive role in our prior evolutionary history (though we will later encounter philosophers who cast doubt on this point), but that is not to concede that intentionality can be identified with the evolutionary fitness of an internal state. Thus, even if all intentional states can be shown to have an adaptive role in the evolutionary history of the organisms that bear them (or shown to be side effects of such directly adaptive states), that does not show that intentionality *just is* playing such a role.

Daniel Dennett, who develops a sophisticated instrumental-

40. Hilary Putnam, *Renewing Philosophy* (Cambridge, Mass.: Harvard University Press, 1995), 30.

ist account of intentionality, attempts to deal with this sort of problem not by finding a place for real intentionality in a universe supposedly governed by purposeless laws and evolutionary natural selection, but by arguing that the notion of intentionality, though very useful for certain tasks, *does not describe an intrinsic property of conscious organisms.* According to Dennett, intentionality is really a heuristic device we use to predict the behavior of certain very complex systems: "The first point to make about intentional systems … is that a particular thing is an intentional system only in relation to the strategies of someone who is trying to explain its behavior."[41] According to Dennett, we attribute intentionality, for example, beliefs, desires, thoughts, to systems as part of a theoretical strategy, a way of predicting what the system will do. The usefulness of intentional concepts does not necessarily imply any intrinsic property had by the system; when we use such concepts we are primarily taking an interpretive stance (what Dennett calls the "intentional stance") toward a system for which talk of "thought" or "desire" is pragmatically useful. Using such examples as computers programmed to play chess at high levels of proficiency, Dennett argues that "a purely physical system can be so complex, and yet so organized, that we find it convenient, explanatory, pragmatically necessary for prediction, to treat it as if it had beliefs and desires and was rational."[42] In other words, the success of an intentional stance for predicting and explaining the behavior of a system does not imply anything inconsistent with materialism regarding that system. We often take the intentional stance toward machines and organisms to which we are not the least tempted to ascribe nonphysical, intrinsically intentional states such as thoughts. That we find talk of thoughts in machines and animals useful doesn't mean they have thoughts.

41. Daniel Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, Mass.: MIT Press, 1981), 3–4. See also Dennett's *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (New York: Simon and Schuster, 1995), 401–27; and *The Intentional Stance* (Cambridge, Mass.: MIT Press, 1989).

42. Dennett, *Brainstorms,* 7–8.

Do we need to say that human beings have thoughts (or other intentional states) as intrinsic features? Dennett, much like in his treatment of qualia we discussed above, appeals to the theoretical sterility of thoughts taken as intrinsic states once we see the explanatory power of the intentional stance. We attribute thoughts to ourselves just when the intentional stance is useful, in much the same way we do so when considering machine and animal behavior. However, in the latter cases we are not inclined to attribute intrinsic, nonphysical intentional states, so then it seems that, at least for theoretical purposes, we can get by without intentionality at all, even in human beings. We don't need *intentionality,* but merely an intentional stance.

My reply to Dennett on this point is likely quite predictable: as in his case against qualia, his attempt to deny that intentionality is an intrinsic feature of conscious organisms presupposes that all that needs to be explained is behavior. There is, however, a fact of intentionality itself. I don't claim that there is a qualitative feel of intentionality, a so-called thought quale, which Dennett is ignoring.[43] Yet, I don't attribute intentionality to myself because that is the only way I can predict my behavior. Rather, I attribute thoughts (as intentional states) to myself because I am aware of my thoughts as being about certain things. Moreover, we might worry that Dennett will run into problems similar to those we raised in reply to the eliminative materialist. Dennett's theory itself is something he believes, a kind of thought. In presenting his theory, Dennett presents us with his thoughts, and in offering the theory he proposes that you and I amend our thoughts inasmuch as we disagree with him; he claims his theory is true. If we attempt to explain away all of these intentional notions, for example, truth, belief, theory, that we must presuppose in our conversation about Dennett's theory as merely heuristic devices, then it is very dif-

43. David Oderberg makes a case that there is such a quale in *Real Essentialism* (New York: Routledge, 2007), 246–47. See also, E. J. Lowe "Naturalism, Imagination, and the Scientific Worldview," in *Turning Images in Philosophy, Science, and Religion,* ed. Charles Taliaferro and Jil Evans, 91–100 (New York: Oxford University Press, 2012).

ficult to make sense of what we are doing when evaluating theories. It would be difficult to take Dennett as proposing his theory as true in the traditional sense of asserting something as corresponding to the way things actually stand in the world, because for him intentional states (such as thoughts and beliefs) that we associate with the traditional notion of truth are not things that represent or correspond to the world in this sense. Of course there are other views of truth that philosophers have proposed that do not so clearly involve intrinsic intentionality, for example, versions of pragmatism that conceive of truth in terms of effectiveness for action and not accurate representation of the world, but once again these notions of truth are frequently subjected to similar accusations of self-referential paradox.[44] Edward Feser makes much the same point when he argues that taking Dennett's intentional stance "is itself a manifestation of intentionality; so we can't coherently suppose that intentionality is a mere artifact of the stance we take toward ourselves."[45] Dennett's approach is basically a "dressed-up" version of eliminativism (though he does not outright deny intentionality, he does deny that it is an intrinsic feature of conscious organisms), which, despite its clever and even ingenious technical virtues, still suffers from some of the vices of eliminativism.

Materialism and Agency

Thus far, our criticisms of materialism (primarily as coupled with the functionalist account of sensation and thought) have focused on only two of the four interesting aspects of human psychological life, sensation and thought, and one of those (sensation) is not even a capacity unique to humans. I will close by

44. This is certainly a very quick rejection of non-correspondence theories of truth, which are a sophisticated and varied lot of philosophical positions held by many very important thinkers, but even a cursory consideration of these issues would take us very far afield of the concerns of the book. For a good introduction to these issues, see Richard L. Kirkham, *Theories of Truth: A Critical Introduction* (Cambridge, Mass.: MIT Press, 1992).

45. Feser, *Philosophy of Mind*, 191.

considering the prospects for a materialist account of the other two aspects of human psychological life, that is, our status as intellectual and moral agents. Not surprisingly, these prospects are rather dim indeed.

Modus ponens is the name of a common pattern of reasoning, which you likely use every day (indeed you have used it a good deal while reading this book), even if you are not explicitly aware of it. *Modus ponens* has the following form:

If P, then Q.
P.
∴ Q.

Any fully rational agent who gives this pattern of reasoning a moment's reflection will likely recognize it as logically valid; there just is no way propositions taking the first two forms can be true without a proposition taking the third form also being true. Suppose we "teach" *modus ponens* to a small child, say my four-year-old son, Jack, who is not yet a fully rational agent, as follows: *Whenever I say, "If P, then Q" and "P," you should say "Q."* Suppose that after extensive training (most likely a process of brainwashing using cookies as a form of bribery) we then pose the following question to Jack: "Given that 'If P then Q' and 'P', what follows?" If properly trained, Jack will say "Q!" Suppose now we pose the same question to you. You will (I hope!) give the same answer.

Notice that both you and Jack will make the same utterance in response to the question (you and Jack are functionally equivalent with respect to *modus ponens*), but you two would not be doing the same thing. Jack isn't really reasoning, but merely doing what he has been told: that is, when you hear the *sounds* "If P, then Q" and "P," make the *sound* "Q." You, on the other hand, understand the logical relations involved and thereby make an inference. Once again, Jack operates only on the basis of rote training or programming, whereas you operate on understanding. In other words, we can explain Jack's saying "Q" entirely in terms

of stimulus and response grounded in programming, whereas in your case we need also to mention an *understanding* of the logical law. We might think of Jack as much like Searle's man in the Chinese room. Jack is not actually an intellectual agent (not yet anyway), but is just parroting the rules of logic without knowing what he is doing.

If functionalism is true, then we all, like Jack, are people in the logic room as we parrot rules of logic on the basis of some prior programming that determines us to make certain responses following certain verbal stimuli, which is a process not essentially different from a computer running on software designed to mimic logic. Computers, however, reason in a particular way not because they grasp the laws of logic, but merely because that is how they happen to be programmed. We might just as easily program a computer to operate according to a fallacy as a valid logical rule. Either way the computer would operate with nearly inexorable consistency. We likewise might just as easily have taught Jack to use a fallacious line of reasoning, and Jack would be none the wiser. In short, functionalism requires us to give up the distinction between actual logical reasoning and merely consistent parroting of reasoning by patterns of stimulus-response behavior.

The objection here is not that, assuming functionalism, *modus ponens* wouldn't be valid; our theories of mind do not determine what is or what is not a law of logic. Rather, the point is that functionalists cannot account for the difference between parroting patterns of valid reasoning in the manner of a pre-rational child and bona fide logical reasoning. The well-trained (yet pre-rational) child and an actual reasoner might be functionally equivalent, but certainly there is a psychological difference between them. Once again, the point here is that logical reasoning poses to the functionalist a problem similar to her difficulty with respect to intentionality as shown by the Chinese Room thought experiment. Suppose we have a man in the logic room, who knows nothing about logic, though he has a detailed manual

that indexes sets of premises (inputs) to valid conclusions (outputs) that would allow him to pass a Turing Test covering basic logic. Does the man in the man in the logic room reason validly? Certainly not; he is merely parroting patters of reasoning he does not understand; for all he "knows," he might just as well have had a manual linking premises to invalid conclusions. The problem for the functionalist is that the man in the logic room and a person who can actually reason validly would realize the same functional states, so there must be a nonfunctional difference that accounts for the distinction between genuine reasoning and parroting logic. Thus, we do not have good reason to conclude that our intellectual agency just is the activity of neurophysiological states that realize certain causal-functions.

It would be unwise for the functionalist to attempt to avoid this problem by suggesting that, as Feser puts it, "we just don't in fact … do *modus ponens,* or carry out any other piece of formal reasoning after all; it only seems like we do because we approximate doing so," because such a suggestion "threatens every argument anyone has ever given for materialism. For if none of us ever really reasons via *modus ponens* or any other valid argument form, then we never reason validly. Every single other argument anyone has given will have been invalid? This materialist response would thus undermine itself!"[46] In other words, if all reasoning is really pseudo-reasoning, for example, we only seem to reason and never actually do so, then any reasoning we might give in favor of functionalism is no exception. Thus, if functionalism is true, then we cannot say that we actually reason our way to it. One could accept functionalism, or any type of materialism, only simply as a matter of blind faith, because on the assumption of the truth of functionalism it is unclear as to what we could even mean by "reason."[47]

46. Ibid., 205.

47. C. S. Lewis famously advanced an argument from reason similar to this in *Miracles,* 17–36. A similar critique can be found in G. K. Chesterton's *Orthodoxy* (San Fran-

A similar problem regarding moral agency and functionalism can be formulated. If functionalism is true, then our actions are not explained by the goods we choose to pursue. We could (and in fact we do) train small children like Jack to parrot real moral agency, before they reach the age of reason. Jack might decide not to steal something because he has been trained not to do so, in much the same way that we could program a robot not to take certain items. That, however, is quite different from a bona fide moral agent deciding not to steal because she understands the goods available to her and her freely choosing to pursue the appropriate good in that situation. Once again, Jack is like a man in the ethical room, who might parrot genuine moral agency, but he lacks the intrinsic psychological states (for example, an understanding of the relevant goods) necessary for such agency. Jack (or anybody in the ethical room) might be functionally equivalent to a genuine moral agent, but there is clearly a difference, but this is not a distinction that the functionalist can readily make. That is, it is unclear how a functionalist can distinguish real moral agency from mere parroting of moral agency. This objection may not be quite as damning as the argument from intellectual agency I present above, because there is nothing strictly self-undermining about claiming that there is no distinction between genuine moral agency and pseudo-moral agency. It does show, however, that a materialist (assuming some version of functionalism is his best bet) must be willing to give up a great deal of what we typically take to be distinctive about ourselves as moral agents. As C. S. Lewis puts it:

---

cisco: Ignatius Press, 1995). See Victor Reppert, *C. S. Lewis's Dangerous Idea: In Defense of the Argument from Reason* (Downers Grove, Ill.: InterVarsity Press, 2003) for a detailed defense of Lewis's version of the argument. More recently, E. J. Lowe has developed a very sophisticated version of a similar argument in *Personal Agency,* 79–91, though I am aware of no direct connection between his argument and Chesterton's and Lewis's. Alvin Plantinga develops an argument against naturalism based on the notion of a reliable cognitive faculty in several places, most recently in in his contribution to *Knowledge of God* and *Where the Conflict Really Lies.* Richard Rorty is among the materialists most explicitly willing to bite the bullet on this issue and accept that his materialism (actually Rorty is an eliminativist) implies that we can no longer distinguish between logical and merely habituated patterns of reasoning.; see Rorty, *Contingency, Irony, and Solidarity.*

But if Naturalism is true, "I ought" is the same sort of statement as "I itch" or "I'm going to be sick." In real life when a man says "I ought" we may reply, "Yes. You're right. That is what you ought to do," or else, "No. I think you're mistaken." But in a world of Naturalists … the only sensible reply would be, "Oh, are you?" All moral judgments would be statements about the speaker's feelings, mistaken by him for something else (the real moral quality of actions) which does not exist.[48]

Michael Ruse, a leading naturalist philosopher, seems to concede Lewis's point when he argues that his naturalism implies that "considered as a rationally justifiable set of claims about an objective something, ethics is illusory. Morality is just an aid to survival and reproduction … and any deeper meaning is illusory."[49] There is nothing logically inconsistent in accepting such a position, but we should at least pause before too quickly accepting any philosophical thesis that would require us to revise our conception of ourselves as moral agents so radically.

Finally, a general problem for the materialist in the case of both intellectual and moral agency is that there just doesn't seem to be a way to account for *normative* facts as somehow supervening on (let alone being identical to) physical states of affairs, at least given the broad assumption of mechanism under which we have been operating throughout our discussion. That is, intellectual and moral reason are normative in the sense that we can evaluate one's performance in these matters against an objective standard; that is, one can be a better or worse theoretical or practical reasoner. We can specify ways that one *should* or *ought* to reason in both intellectual and moral matters, but a mechanistically con-

48. Lewis, *Miracles,* 51. Notice that Lewis mentions "naturalism," but we are for now still assuming that materialism is the naturalist's only option. We will reconsider this assumption in the following chapter.

49. Michael Ruse, "Evolutionary Theory and Christian Ethics, " in *The Darwinian Paradigm* (London: Routledge, 1989), 268–69. This passage was brought to my attention by J. P. Moreland's discussion of an argument against naturalism based on moral agency in *The Recalcitrant Imago Dei: Human Persons and the Failure of Naturalism* (London: SCM Press, 2009), 243–64. Not all naturalists are materialists, and there are materialists who are not naturalists. We have been considering only materialist positions in the service of naturalism, so it is fair to conclude that Ruse's comments apply to our discussion.

strued physical world is devoid of any *should* or *ought* relations. Thus, if matter is all that there is, it is implausible to suppose that we can account for moral and intellectual agency because of "the difficulty of seeing how one could possibly give materialistic or naturalistic explanations for putative ontological supervenience-relations between natural properties and facts and putative normative properties. Objective moral values do not appear to be part of the natural."[50] Moreover, because "*epistemic warrant* is a normative concept too," we should likewise conclude that intellectual agency is not likely something we can square with materialism.[51] Of course materialists are at work in attempting to deal with these problems, but it is far from clear as to how they can solve them without denying the objectivity of our normative notions, which once again raises the specter of paradox. As Terence Horgan puts it, given the problems of normativity, "it is not easy formulating a metaphysical position that meets the demands of a material world; there is still a lot of philosophical work to do."[52] With these seeming Herculean labors left as yet incomplete, one may well be within her rights to look beyond materialism in the philosophy of mind.

Throughout the foregoing chapter we have discussed the most common objections to materialism, stemming from sensations, thoughts, and agency, along with the standard replies by materialists. A recurring pattern has emerged from this discussion, namely that materialism, whether it treats thoughts or sensations, always seems to leave out or eliminate central psychological phenomena. Of course the debate regarding materialism is ongoing, and we have treated many important issues in a merely cursory manner. Searle, nevertheless, seems correct to conclude that "what we find

50. Horgan, "From Supervenience to Superdupervenience," 159.
51. Ibid.
52. Ibid. Horgan suggests some *irrealist* strategies for giving a materialist account of normativity in this paper.

in the history of materialism is a recurring tension between the urge to give an account of reality that leaves out any reference to the special features of the mental, such as consciousness and subjectivity, and at the same time account for our 'intuitions' about the mind," and there is little hope of a materialist solution to this tension.[53] The question for us is whether the failure of materialism spells the end of naturalism. We will see in the next chapter that there are very sophisticated naturalists who believe that naturalism can get by very well without materialism.[54]

53. Searle, *The Rediscovery of the Mind*, 52.

54. While completing this chapter, I have had the benefit of a great many comments and criticisms from an anonymous reviewer that have been particularly helpful. Whether the final version is satisfactory to him or her, I do not know. In any event, this chapter is much the better because of the care he or she took in commenting on the manuscript.

CHAPTER 6

EMERGENTISM AND NATURALISM

In the last two chapters, we considered the prospect of constructing a materialist version of naturalism, with particular regard to the mind-body problem. We found that materialists typically attempt either to eliminate psychological states (sensations and thoughts) entirely or identify them with straightforwardly physical entities, usually as features of the central nervous system. The most extreme of these materialist endeavors is *eliminative materialism* (the doctrine that mental events and substances should be rejected entirely as postulates of a now thoroughly refuted quasi-scientific theory) and *logical behaviorism* (the doctrine that all our words seemingly referring to psychological states can be translated into language referring only to overt behavior or dispositions for such behavior), but we found quickly that these doctrines have such absurd consequences that they do not enjoy wide currency even among contemporary materialists. We then turned to a somewhat less extreme sort of materialism, that is, *the identity theory,* which admits the reality of sensations and thoughts, but

claims that such states are identical to more fundamental neurophysiological phenomena. Many philosophers believe that the identity theory's central claim that *types* of psychological states can be identified with *types* of neurophysiological states is far too ambitious, and they subsequently propose a *nonreductive materialism* that claims only that each *token* of a psychological state is in fact a *token* of a neurophysiological state, while types of psychological states and types of neurophysiological are not identical. We then found that *functionalism,* according to which psychological states are defined in terms of causal roles linking environmental inputs with behavioral outputs, is likely the best option for the materialist to develop a nonreductive position, even though it is not essentially a materialist doctrine. The shared difficulty for all of these versions of materialism is that they force us to deny what seems obvious: there are such entities as sensations and thoughts that have qualitative and intentional features that simply don't seem to be physical. Moreover, we discussed the fact that materialism is seemingly unable to account for the fact that human beings are both intellectual and moral agents, which has counterintuitive, and maybe even paradoxical, results.

In light of these seemingly intractable problems, some philosophers attempt to develop nonmaterialist versions of naturalism, wherein either psychological substances or properties are not strictly physical entities though they emerge through the interactions among strictly physical entities. What follows in this chapter is a discussion of naturalistic emergentism and some of the standard criticisms of this view. In this section I will introduce naturalistic emergentism by discussing a position defended by John Searle, which he calls biological naturalism. Emergentism, however, is defended by various contemporary and historically significant philosophers, so it would be unfortunate to associate the view with Searle alone. I have, nevertheless, chosen Searle's presentation, because among recent emergentists he is the philosopher most willing to question the fundamental assumptions

unpinning much of contemporary philosophy of mind, a point that will be very important throughout the remaining chapters of this book.[1]

Emergence

We have seen Searle claim in various ways that the philosophy of mind is at a stubborn impasse, wherein we want to affirm what seems like a contradiction: "On the one hand we accept a view that seems overwhelmingly convincing—the universe is material—but that seems inconsistent with another view that we cannot give up—minds exist."[2] The problem is that while the common

1. For recent examples of emergentism, in addition to Searle's work, see Timothy O'Connor, *Persons and Causes* (New York: Oxford University Press, 2000), 10–126, and "Emergent Properties," *American Philosophical Quarterly,* 31 (April 1994): 91–104; Nancy Murphy, *Bodies and Souls, or Spirited Bodies* (Cambridge: Cambridge University Press, 2006); and Tooley's contribution to Plantinga and Tooley, *Knowledge of God,* 189–205. Neither O'Connor nor Murphy, though they are emergentists about human minds, can be characterized as *naturalists* as I have defined the term in this book (they are both theists!), though they give accounts of the relationship between mind and body that could in principle be accepted by a naturalist of a very broad sort. Historical proponents of this sort of view can also be found: One of Socrates's interlocutors in the *Phaedo* seems to defend something that can be interpreted as a version of emergentism arguing that the soul is "the harmony of the body"; see Plato, *Phaedo,* translated by G. M. A. Grube (Indianapolis, Ind.: Hackett Publishing, 1978). John Locke famously entertains (though he does not endorse) the possibility of something like emergentism in his *Essay Concerning Human Understanding,* ed. P. Nidditch (Oxford: Claredon Press, 1975), 39; Hume seems to take such a position in his essay "On Immortality," in *Dialogues on Natural Religions, with Other Essays* (Indianapolis, Ind.: Hackett Publishing, 1980); and C. D. Broad defends a very influential version of emergentism in *Mind and Its Place in Nature* (London: Routledge and Kegan Paul, 1925). Without endorsing the position, P. E. Meehl and Wilfrid Sellars consider emergence in detail in "The Concept of Emergence," in *Minnesota Studies in the Philosophy of Science,* ed. Herbert Feigl and Michael Scriven, vol. 1: *The Foundations of Science and the Concepts of Psychology and Psychoanalysis,* 239–52 (Minneapolis: University of Minnesota Press, 1956). For more recent versions of emergent substance dualism, see the following: Hasker, *The Emergent Self;* Swinburne, *The Evolution of the Soul;* Popper and Eccles, *The Self and Its Brain;* Unger, *All the Power in the World;* Lowe, *Personal Agency;* and Dean Zimmerman, "From Experience to Experience," in *The Soul Hypothesis: Investigations into the Existence of the Soul,* ed. Mark C. Baker and Stewart Goetz, 168–201 (New York: Continuum, 2011). See O'Connor, *Persons and Causes,* 110–11, for several more references to thinkers who defended versions of emergence in the nineteenth and twentieth centuries. Searle's view is not always completely transparent; for example, he sometimes refers to his position as "causal supervenience" (see Searle, *The Rediscovery of the Mind,* 124), but in what follows it will be clear that what he has in mind is quite distinct from the supervenience views we have discussed in the last chapter.

2. Searle, *Mind: A Brief Introduction,* 75.

mechanistic understandings of nature, bolstered by the success of natural science, lead us to presume we can account for ourselves as just one more type of physical object, at the same time it appears that we can do so only at the expense of the fact that we have minds. Try as we might, the philosophy of mind keeps slipping back into some form of dualism or a materialism that seems "to leave something out."

Searle is a naturalist who thinks that dualism is a disaster, for many of the reasons we considered in chapter 3. Although he claims to have the solution to the mind-body problem, he does not resolve the issue in favor of materialism. Rather than offering yet another supposed reduction of the mental to the physical or constructing a just-so story as to why we don't really need psychological notions to make sense of ourselves, Searle attempts to *dissolve* the issue. The mind-body problem, according to Searle, really is no problem at all. It is what philosophers call a *pseudo-problem,* or the appearance of a problem we might fall into, given certain wrong-headed assumptions. Once we have disabused ourselves of these misguided assumptions, we can see our way clear of the mind-body problem entirely. I will begin our discussion of Searle by taking up each of the four assumptions he believes have set us on the wrong track, along with his reasons for thinking that they are philosophically dubious in their own right (independently of their consequences in the philosophy of mind). We will then discuss Searle's emergentist theory of mind, *biological naturalism.*

(a) The first and foremost of the mistaken assumptions motivating the mind-body problem is the belief that "'mental' and 'physical' name mutually exclusive ontological categories," such that something's being mental implies it is nonphysical and its being physical implies that it is nonmental.[3] It is not surprising, as we discussed in chapter 2, that if one makes such an assumption, then it seems that we are forced into a dichotomy between

3. Ibid., 76.

dualism and materialism: either mind exists and is therefore distinct from the body, or it is just the same thing as the body, which means it is not psychological at all; that is, there isn't *really* a mind. Searle is puzzled as to why anybody would really think this assumption is true in light of our current understanding of physical reality. If we accept Descartes's understanding of matter, one might make this assumption, since he defined matter in terms of spatial extension, and sensations and thoughts are certainly not extended. The problem for that view, however, is that we now know that there are physical things, for example, electrons, that don't fit so neatly into the early modern physics, so our notion of what counts as physical has expanded since the seventeenth century. As Galen Strawson puts it, the supposed dichotomy between mind and matter looked plausible "in Descartes time, for classical mechanistic materialism … was then the dominant view.… But the strict mechanist understanding of the physical world was fatally undermined" by subsequent progress in physics.[4] Searle recommends that we define the physical as follows:

> First, real physical phenomena are located in space-time. (Thus electrons are physical and numbers are not.) Second, their features and behaviors are causally explainable by microphysics. (Thus, solidity and liquidity meet this test. Ghosts, if they existed, would not.) Third, where real, physical phenomena function causally. (Thus solidity is a real physical phenomenon. Rainbows … are not real physical arches in the sky. They do not cause anything.)[5]

In other words, the mark of the physical is spatial location, explanation ultimately in terms of the physics of fundamental particles, and causal power to affect other physical entities. If we define "physical" in terms of extension alone, then we would seem to exclude psychological states, and we would therefore have to

4. Galen Strawson, *Real Materialism and Other Essays,* 40. When Searle argues that modern physics calls the mechanist theory of matter into question, he has in mind contemporary particle physics, whereas Strawson believes that mechanism is a dead letter even with Newton's *Principia* in 1687. See Strawson's discussion in *Real Materialism* for details.

5. Searle, *Mind: A Brief Introduction,* 83.

declare them to be *mental* (nonphysical). If we define "physical" in the way Searle outlines above, which is more in line with what modern physics has revealed, then there is no reason to exclude the psychological states from the physical world: "There is no reason why a physical system such as a human or animal organism should not have states that are qualitative, subjective, and intentional."[6]

(b) Philosophers of mind typically assume that "the notion of reduction … is clear, unambiguous, and unproblematic," but Searle argues that we need to distinguish between two different types of reduction, "ontological reduction," which seems to include both type-identity and token-identity (supervenience), as we have understood these notions thus far, and more modest "causal reduction."[7] To say that an individual, *x,* is *ontologically reducible* to *y* is to say that, though *x* is real, *x* is identical to some more fundamental entity (or set of entities), *y.* For example, the projector hanging in room 401 of St. Benedict's Hall is strongly reducible to the set of parts that compose it; that is, it just is those parts.[8] To say that *x* is *weakly* or *causally reducible* to *y* is to say that *x* is not identical to *y,* but that all of the intrinsic facts about *x* are explained by (or caused by) facts about *y.*[9] In other

6. Ibid.

7. Ibid., 76–86. It seems that Searle is including both type-identity and token-identity (supervenience) under the moniker "ontological reduction." At one point Searle does speak of his weak or causal notion of reduction in terms of supervenience, but he clearly distinguishes it as "causal supervenience" in contrast to what he calls "constitutive supervenience," which seems to be what most philosophers of mind mean by "supervenience." See Searle, *The Rediscovery of the Mind,* 124. I discuss Searle's position in terms of supervenience in "Realism, Nominalism, and Biological Naturalism," *International Philosophical Quarterly* 51, no. 1 (2011): 85–102.

8. This claim is actually ambiguous. One could mean that (a) the projector is always identical to the set of parts that compose it right now, such that if just one of these parts were replaced it would be a new entity; or (b) the projector is composed by some collection of parts or the other at any time it exists, though it could be identical to different collections at different times. For our purposes, I don't believe I need to take a stand on either of these interpretations of identity or strong reduction as applied to objects like projectors. There is a very large literature or this issue, but for a highly influential introduction see Roderick Chisholm, *Person and Object* (Chicago: Open Court, 1976).

9. Notice that *x* could be causally reducible to *y* and yet there could be *extrinsic* or

words, there is nothing about *x* that isn't dependent on *y*. For example, the image projected on the screen by the projector in room 401 is not ontologically reducible to the projector, but it is causally reducible to the projector, because features of the projector determine all of its features. The crucial difference is that in cases of ontological (strong) reduction, the thing reduced is said to be identical (in some sense) to the more fundamental entity, whereas in cases of causal (weak) reduction, the thing reduced is said to be distinct but completely dependent on the more fundamental entity for all of its features, though it remains a distinct entity. We might say, though this is not how Searle puts it, in cases of strong reduction there is at least relation of token-identity (and maybe type-identity), whereas in weak reduction there is no claim of identity in any sense, just causal dependence. As we shall see, Searle recommends that we take the relation between mind and body in terms of an analogy with the relation not between the projector and its parts, but between the image and the projector; psychological states are causally reducible to processes in the brain, even though they are not ontologically reducible.

The notion of causal reduction is closely associated with another key concept in Searle's philosophy of mind, *emergent properties* or *system-level features.* One of Searle's favorite examples is the solidity of the table on which my computer is currently resting. This seemingly mundane fact is interesting, because none of the particles that ultimately compose the table are themselves solid in the same sense in which the table is solid; certainly, none of these particles individually, in a small group, or improperly arranged possess the solidity necessary to prevent the computer from crashing to the floor. Thus we can't be strong reductionists about solidity. We nevertheless are not going to adopt solidity-particle dualism. The solidity is an effect of the arrangement of the particles—it is causally reducible to the arranged particles. Solidity is

relational facts about *x* that are not explained about *y*. For example, the fact that somebody is thinking about *x* might have nothing to do with *y*.

not a separate entity, but an emergent feature that arises from the proper arrangement of physical particles. Searle rightly points out that nature is full of such relations, for example, liquidity–H_2O, digestion–stomach cells, magnetism–charged iron atoms. Solidity, liquidity, digestion, and magnetism are systems-features or emergent properties causally reducible to entities composing the appropriate system, that is, subatomic particles, H_2O molecules, stomach cells, and charged iron atoms. Even though the parts of these systems lack these features on their own (e.g., an H_2O molecule isn't wet), certain novel emergent properties arise once the system is properly organized.

(c) With the notions of emergence and causal reduction in mind, we can understand Searle's rejection of the third typical assumption lying behind the mind-body problem. Many modern philosophers also assume that "causation is always a relation between discrete events ordered in time, where the cause precedes the effect."[10] In other words, it is common to take causation exclusively in terms captured by billiard ball examples; for example, the eight ball's motion was caused by the impact of the cue ball. It is not difficult to see how one might be led to dualism in this way, because the only model for mind-body interaction (which we supposedly must accept on pain of epiphenomenalism) implies that the mind and body are discrete, independent entities. Searle, however, points out that causation need not occur exclusively between discrete objects like billiard balls, and this should be readily apparent from the examples of emergent properties that we discussed above; solidity, liquidity, and so on are effects of the parts of certain systems of particles, even though these emergent features are not in any way independent or discrete.

Nondiscrete causation is not just *bottom-up* (i.e., moving from microstructure to a system-level feature), but also *top-down* (i.e., moving from the system-level feature to the particles composing

10. Searle, *Mind: A Brief Introduction*, 77.

the microstructure). For example: "Consider a wheel rolling down hill. The wheel is entirely made of molecules. The behavior of the molecules causes the higher-level, or system feature of solidity. Notice that the solidity affects behavior of the individual molecules. The trajectory of each molecule is affected by the behavior of the entire wheel."[11] The solidity of the wheel emerges from the arrangement of the particles, but that very system-feature can likewise affect the motion of the particles from which it emerges. Consider also the way a magnetic field, which emerges from iron atoms, can cause those very atoms to move when another magnet with the same polarity is brought into its field.[12] These examples show that even in cases in which one entity is causally reducible to the other, for example, the solidity of the wheel and the particles composing the wheel, there can be bona fide causal relations in both directions, bottom-up and top-down.

(d) These considerations bring us to the fourth assumption Searle advises us to jettison: "Identity … is assumed to be unproblematic. Everything is identical with itself and not with anything else. Paradigms of identity are object identities and identities of composition."[13] By an identity of composition, Searle means something like the identity between the chair I am sitting on and the parts that compose it, and by an object identity Searle has in mind something like the fact that the Morning Star is the Evening Star. Modern philosophers typically assume that if we say "sensations and thoughts just are features of the brain," then we must mean that there is something in the brain to which sensations and thoughts have either an object or composition identity relation. As we have seen in the last chapter, it is difficult to conceive of anything in the brain identical to sensations and

11. John Searle, *Freedom and Neurobiology: Reflections on Free Will, Language, and Political Power* (New York: Columbia University Press, 2007), 48. Searle himself adopts this example from another emergentist; see Roger Sperry, "In Defense of Mentalism and Emergent Interaction," *Journal of Mind and Behavior* 12, no. 2 (Spring 1991): 230.

12. William Hasker introduces this example in *The Emergent Self*, 190–91.

13. Searle, *Mind: A Brief Introduction*, 77.

thoughts, so this way of thinking lends itself to either dualism or eliminativism. Returning to the example of the wheel, Searle believes it is clear that there is a sense in which an entity can be a feature realized in or emergent from something, without being strictly identical to any of its parts:

> The wheel consists entirely of molecules. So when we say that solidity functions causally in the behavior of the wheel and in the behavior of the individual molecules that compose the wheel, we are not saying that the solidity is something in addition to the molecules; rather, it is just the condition that the molecules are in. But the feature of solidity is nonetheless a real feature, and it has real causal effects.[14]

The solidity of the wheel is neither objectually nor compositionally identical to the arrangement of the molecules that compose the wheel, but it is likewise not an independent entity in its own right. Rather, solidity is something that is causally reducible to the arrangement of particles, a system-level feature realized in the microstructure of the wheel. Although we cannot point out any particular part of the wheel that has the solidity (all the ultimate parts lack this characteristic), the solidity is a feature of the wheel. Even if we conclude that the mind is not a discrete entity in its own right, we should not presume that we must find strict identity between the mind and some part of the brain.

Having rethought these assumptions along the lines that Searle recommends, we find it fairly easy to see the handwriting on the wall regarding his dissolution of the mind-body problem, what he calls *biological naturalism:* "Mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain. To distinguish this view from many others in the field, I call it "biological naturalism." Mental events and processes are as much part of our biological natural history as digestion, mitosis, meiosis, or enzyme secretion."[15] In short, for biological naturalism, conscious states (sensations and thoughts) are

14. Searle, *Freedom and Neurobiology,* 49.
15. Searle, *The Rediscovery of the Mind,* 1.

system-level features of neurophysiological processes in the central nervous system, which is to say that they are *causally reducible* to and emergent from those very same neurophysiological processes. The biological naturalist embraces the reality of psychological phenomena, and makes no attempt to eliminate them or identify them with neurophysiological events. This allowance frees the biological naturalist from the burden of many of the thorniest problems in the philosophy of mind, because she need not make strong claims about the reduction, supervenience, or elimination of psychological phenomena that fly in the face of common experience. Psychological phenomena may be qualitatively different from physical phenomena, so the biological naturalist might concede, but they are nevertheless system-level features of the brain caused by lower-level, neurological features of that same system. Just as we can understand the lower-level physical structure of the particles composing a table as the cause of its solidity, while neither invoking particle-solidity dualism nor explaining away solidity, we can likewise understand that "lower-level processes in the brain cause my present state of consciousness, but that state is not a separate entity from my brain; rather it is just a feature of my brain at the present time."[16]

Searle does not believe he has given us an ad hoc solution, because the sort of relation he proposes between the mind and brain is found in many other natural systems, as we have noted above. Of course it is a bit strange, at first blush, to claim that even though the parts of the brain are essentially nonpsychological, the brain has psychological features. Searle points out repeatedly, however, that we are well acquainted with systems that have features that are not present in any of the components of the system. Since there is no general mystery about systems features, so there should be no such mystery regarding conscious states.

16. John Searle, *The Mystery of Consciousness* (New York: New York Review Books, 1997), 8.

Searle's claim here is that conscious states just are features of the physical world no different from digestion, magnetism, liquidity, and so on. Like these other emergent properties, sensations and thoughts are located in time-space (i.e., they occur in our heads), their existence is explained by indubitably physical objects (i.e., they are causally reducible to neurophysiological processes), and they have physical effects (i.e., they exhibit downward causation on the brain). Of course, conscious states are not strictly identical to their neurophysiological bases, but all of their powers are extensions or effects of the powers of their neuronal base. This point is crucial for Searle, because it "shows that we are not talking about two independent things, consciousness and neuronal processes."[17] Thus, Searle believes that his position avoids both substance and property dualism, because sensations and thoughts are physical in the same sense that digestion, solidity, liquidity, and so on, are all physical phenomena, and just like these other physical, emergent features, conscious states in no way operate independently of their foundations.[18]

Searle also believes that his position avoids materialism, because he takes consciousness as "irreducibly qualitative, subjective, first-personal, airy-fairy, and touchy-feely," even though it is a system-feature of the brain.[19] What allows Searle to avoid materialism is his recognition of the irreducible first-person aspect of psychological states; we can't say that psychological states *just are* neurological processes (in the sense of being strictly identical to some part of the brain) because no first-person perspective *just is* a third-person phenomenon. Thus, he avoids materialism by

17. Searle, *Mind: A Brief Introduction,* 89.

18. Timothy O'Connor, however, seems willing to think of his version of emergentism in terms of property dualism, or at least a dualism of powers; see O'Connor, *Persons and Objects,* 11–117. Others, including Hasker, Unger, Swinburne, Zimmerman, and Lowe, who are impressed with some version or other of the unity of consciousness argument or the argument from the problem of the many we discussed in chapter 2, cast emergentism in terms of substance dualism in order to account for the simplicity and/or unity of the subject of sensation, thought, and agency; see *The Emergent Self, All the Power in the World, The Evolution of the Soul,* "From Experience to Experiencer," and *Personal Agency.*

19. Searle, *Mind: A Brief Introduction,* 88.

abandoning the program of ontological reduction, but he maintains that "consciousness is causally reducible. . . . .It is part of the ordinary physical world and is not something over and above it."[20] Searle eschews materialism, but he holds fast to naturalism.

Take the case of Patrick's thought that "Martha is beautiful," which is ontologically distinct from any lower-level, neurophysiological features of Patrick's brain. Even though materialism is not a promising option, we need not back-slide into dualism, because Patrick's thought is a higher-level feature of his brain caused by its lower-level features. Patrick's brain (or the state of some part of it) causes his belief, and there is nothing about his thinking that calls for any explanation beyond the neurological facts about Patrick anymore than the liquidity of water requires an explanation in addition to the facts about H_2O molecules. In both cases we have higher-level features of a system not identical to, but causally dependent on, lower-level features of that same system.

If Searle is correct, then the mind-body problem should no longer vex us. The mind is real, though physical in exactly the way liquidity and solidity are physical. The question about the mind is no longer "What is it?" but "How does it work?" Since the mind is a system-feature of the brain, the work left to be done is to consider how brains produce consciousness: what are the neurophysiological processes from which conscious states emerge, and under what evolutionary conditions have these processes come to be standard features of certain species of animals? In other words, our concerns when we wonder about the mind are now questions for neuroscience, which tells us how the brain works, and evolutionary biology and psychology, which tell us why the brain has come to work in some particular way through our natural history. Either way, questions about the mind become scientific matters, not occasions for metaphysical speculation. Re-

20. Ibid.

manding the mind to the empirical sciences, however, does not necessarily entail that we must lose ourselves as conscious and rational agents, because Searle believes he has avoided even the slightest hint of the difficulties that lead many philosophers to reject materialism.

Problems for Emergence

If Searle is indeed correct, philosophers of mind should be preparing to close up shop, because the mind-body problem is about to go away. Impressive as biological naturalism may be, Searle's announcement of the dissolution of the mind-body problem is more than a bit hasty, and in the remaining three sections of this chapter we will discuss lines of criticism typically leveled against emergentism. Throughout this discussion I will continue to take Searle's biological naturalism as representative of emergentism in general, though at some points we will discuss how other emergentists have addressed these criticisms.

Searle claims that nature is replete with cases of emergence, for example, solidity, liquidity, magnetism, and the like. One might wonder whether these are in fact bona fide cases of emergence, and if they are emergent, one might likewise wonder whether emergent properties are really all that interesting. Suppose we have a collection of matches arranged pentagonally. In this case, the whole of the collection has a property lacked entirely by its members. Should we then say that the pentagonal arrangement is an emergent feature? If we answer affirmatively, "then emergent properties are nothing special. You could put this by saying that that pentagon is 'nothing over and above' its constituent matchsticks suitably arranged. The property being pentagonal 'comes for free.' It is not something that must be added to the world in addition to the matchsticks and their arrangements."[21] Indeed, if that's all there is to emergence, it would be quite difficult to dis-

21. John Heil, ed., *The Philosophy of Mind: A Guide and Anthology* (New York: Oxford University Press, 2004), 678.

tinguish it from some version of supervenience, and in some of Searle's examples we might do well to take the relation as just that. We will return to this concern a bit later, but for now notice that all of Searle's examples (to his satisfaction at least) involve novel causal powers; that is, in each case there are significant causal powers possessed by the whole system that are not possessed by its parts. None of the subatomic particles have that power to prevent the mug from falling to the floor, none of the H_2O molecules alone have the powers of liquidity, none of the stomach cells alone has the power to dissolve a steak, and so on. Thus, Searle might be able to claim plausibly that all of his examples are interestingly emergent, because they involve novel causal powers as system-level features.

Let's then take Searle's examples as legitimate cases of emergence, at least for the moment. If we attend closely to these examples, we will find that the emergence relationship, rather than being utterly mysterious, is quite intelligible, and this intelligibility comes in two kinds. First, consider a simple box composed of five sheets of plywood and some wood screws.[22] Suppose that once constructed this box is used as a storage container for a basketball. Neither the sheets of plywood nor the screws have the power of containing a basketball independently of their composing a box, so the power of containment is an emergent feature. Notice, however, containment is not a mysterious brute fact that tags along with the parts of the box. We can understand *why* such a feature would emerge from a certain arrangement of screws and plywood. Indeed, given what we know about the essential properties of wood and steel, we can see why the power of containment is necessitated by a certain structural arrangement of wood screws and plywood. The point here is that in cases like the emergence of the power of containment, we can give an explanation of

22. I have adapted this example from Goetz and Taliaferro, *Naturalism,* 75, who cite Marvin Minsky, *The Society of Mind* (New York: Simon and Schuster, 1985), 28, as its original source.

the emergent property as a structural feature of the parts of the system; that is, we can see that given the properties of the parts, in addition to their arrangement, the emergent property had to occur. Searle's examples of the emergence of solidity from the structure of the particles that compose the table and the emergence of liquidity from large quantities of H_2O molecules are seemingly cases of this sort of *structural emergence.*

Consider the example of a choir of small children. Each child sings quite softly, but the choir they compose is quite loud. This too might be a case of emergence, but it isn't a result of the structural relation among the children or their voices; a similar phenomenon could occur through a cacophony of soft voices as when a room full of people engaged in relatively quiet conversations becomes very loud. In these cases the accumulation of sufficiently many soft voices eventually crosses a threshold beyond which a great volume of noise is achieved. Maybe this volume even outstrips what we would expect just by "adding" the voices together, but it is entirely dependent on the soft voices of the children. Here too, however, the emergence isn't mysterious. Even though none of the children is individually loud, each of them makes *some* noise; *they are all a little bit loud.* The property of being loud is contained to a lesser degree in each of the parts of the system. It is then no mystery that when we conglomerate enough of these *proto-louds* we get a bona fide loud phenomenon. Searle's examples of digestion emerging from stomach cells and the example of a magnetic field emerging from the charged iron atoms are examples of this sort of emergence; that is, each stomach cell secretes some small amount of acid capable of digesting a little bit, and each iron atom is a little bit magnetic. Even in cases in which the entities composing the emergence base do not literally have a proto-version of the emergent feature, we would find emergence plausible if we nevertheless have reason to believe that these entities have a natural potential for these features. Thus, in these cases emergence is not at all mysterious,

because there is only a quantitative difference (not a qualitative difference) between the emergent feature and a key feature (whether actual or potential) of the parts of the system; that is, the emergence base and the emergent feature are not fundamentally or qualitatively different kinds of phenomena. Let's call this relation *quantitative emergence.*

Consider now the following argument:

(1) If psychological states emerge from neurophysiological states, then they do so by either (a) structural emergence (psychological states are structural features of the brain), or (b) quantitative emergence (the parts composing the nervous system have proto-consciousness). (premise)

(2) Condition (a) is not satisfied, because psychological states cannot be accounted for in terms of structural complexity. (assumption)

(3) Condition (b) is not satisfied, because the elements of the central nervous system do not possess a proto-version of consciousness. (premise)

Therefore:

(4) Psychological states do not emerge from neurophysiological states. (from 1–3)

The case for premise (2). For the moment, we will set premise (1) aside, though we will scrutinize it in great detail below. Why might somebody think that (2) is true? Goetz and Taliaferro argue that something like (2) is true because psychological states are metaphysically simple; sensations and thoughts don't have parts in any sense, so they cannot be composed of properly arranged, simpler things like the parts of the brain.[23] Moreover, some emergentists argue that it is not just nonphysical, psychological properties or powers that emerge from the complexity of the nervous system, but a nonphysical substance; the mind is a nonphysical substance, distinct from the body, which emerges from the body. These emergent, substance dualists often defend their position in part by ap-

23. See Goetz and Taliaferro, *Naturalism,* 72–82.

peal to the unity of consciousness argument or some version of the argument from the problem of the many we discussed in chapter 2.[24] Thus, the emergent, substance dualist typically claims that the emergent substance is metaphysically simple. Goetz and Taliaferro's point is that one simply cannot derive metaphysical simplicity from complexity. At least as a matter of common sense, nobody doubts that a complex thing can come to be through the combination of distinct parts, but the issue here is whether something utterly simple could come to be by combination. It seems that the various parts would have to make distinct contributions to the emergent entity, so there would be some division or distinction within such a thing; that is, the aspect or part of the simple entity to which constituent *x* contributes would be distinguishable from the aspect or part to which constituent *y* contributes. Whether this sort of internal distinction is contrary to the sort of simplicity (supposedly) essential to psychological states or substances is difficult to answer, and we won't go any further down the path of addressing the metaphysics of simplicity here.[25] Timothy O'Connor, who defends a version of emergentism (though not the substance dualist variety) would actually be happy to concede something like (2), because he does not think that emergent properties are merely structural features of their emergence base: "an emergent property is nonstructural," and is "'simple,' or partless."[26] As some emergentists themselves grant that however emergence works, it is not a matter of structural complexity, we should take Goetz and Taliaferro's way of defending something akin to premise (2) as plausible, even if not decisive.

24. Seen note 18 for references.

25. If the division is only a distinction among the properties of a simple substance, then there is no problem here for the emergent, substance dualist; there is no reason to think that a simple, psychological substance would have only one property. Depending on whether one is willing to countenance higher-order properties (properties of properties), this consideration may count in favor of the emergent, property dualist.

26. O'Connor, *Persons and Causes,* 111. Notice that O'Connor does not believe emergence need satisfy either (2) or (3), because he argues against (1) by defending *brute emergence,* which we will discuss later in this section.

The Case for Premise (3). Strawson raises consideration in favor of something like premise (3):

> You can get liquidity from non-liquid as easily as you can get a cricket team from eleven things that are not cricket teams. [In an ideal physics] it would have to be just as plain how you get experiential phenomena from wholly non-experiential phenomena. But this is what boggles the human mind. We have … no difficulty with the idea that liquid phenomena (which are wholly [physical] phenomena) are emergent properties of non-liquid phenomena (which are wholly [physical] phenomena). But when we return to the case of experience, and look for an analogy … it seems that we can't make do with things like liquidity, where we move wholly within a completely conceptually homogeneous (non-heterogeneous) set of notions. We need an analogy on a wholly different scale if we are to get any imaginative grip on the supposed moved from the non-experiential to the experiential.[27]

Strawson's point is that the emergentist such as Searle would have us think of the relations between the mind and the brain in terms of an analogy with the relation between liquidity and H_2O molecules; just as liquidity emerges from nonliquid entities, so consciousness emerges from nonconscious entities. We can see how the emergence of liquidity works just as readily as we can see how a group of players causes the "emergence" of a sports team, but this is not the case with respect to psychological states and the brain. First, as Strawson points out, in the players-team and liquidity-H_2O cases, we are dealing with straightforwardly *homogeneous types of phenomena;* that is, these phenomena are straightforwardly species of the same fundamental kind of phenomena. The players are not literally little teams, but they are clearly the kinds of things that can be team members under certain conditions. Nobody would expect that teams and players were somehow utterly unrelated kinds of things, nor would it occur to anybody that liquidity is a nonphysical phenomenon. We would say the same about the emergence of the sound of the choir or the power of magnetism and digestion; in all of these cases the emer-

27. Strawson, *Real Materialism,* 63.

gent entity is of the same fundamental type, that is, straightforwardly physical, as the emergence base. We can then see a natural potential for such an emergence relation. This is not the case with psychological states, for by Searle's own admission such states have a property, that is, first-person perspective, that is utterly foreign to all of our concepts of ordinary physical objects; consciousness and ordinary physical objects certainly seem to be the most *heterogeneous types of phenomena* possible (they are fundamentally differ kinds of phenomena): "The experiential/nonexperiential [conscious/nonconscious] divide, assuming that it exists at all, is the most fundamental divide in nature,"[28] so to say that psychological states can emerge from neurophysiological states based on an analogies proposed by Searle is rather strained. In none of the supposed cases of emergence are we dealing with radically heterogeneous phenomena, but that is exactly what we find in the consciousness/nonconsciousness case.

One way philosophers have circumvented (3), is to claim that matter at its most fundamental level is proto-conscious. We call this view panpsychism, which is to say that consciousness is present in the basic physical particles that make up the universe. According to the panpsychist, consciousness is among the fundamental properties of matter, like mass and charge; it is a feature of all physical entities, and can be explained by nothing more basic. The quantitative emergence of psychological states is plausible, because the parts of neurophysiological systems are proto-conscious, as are the particles that compose everything else. That is, large-scale consciousness emerges from an organization of small-scale consciousnesses, in just the same way as a great volume of sound emerges from a collection of small volumes of sound in the children's choir. Consciousness in us (sensations and thoughts) has a perfectly natural explanation, because proto-consciousness is one of the fundamental physical features of the

28. Ibid., 65.

natural world. For this reason, panpsychism is a view sometimes defended by naturalist philosophers of mind.[29]

Searle points out that panpsychism not only is inherently implausible, but also faces a serious problem in trying to make sense of the unity of consciousness: I am one conscious substance, but according to the panpsychist, I am composed of billions of conscious substances. How do we get a single subject out of legions of "little" subjects? Leaving that objection aside, the inherent implausibility of panpsychism alone is enough to make us wonder why anybody would accept such a view. It seems, to say the least, a bit ad hoc to posit consciousness as a fundamental feature of the physical universe, unless one has a prior commitment to naturalism, which once again begs the question we set out to answer. There is nothing I can point to (assuming the unity of consciousness issue could be addressed without more emergence hand-waving) as utterly incoherent about panpsychism, just as we discussed in chapter 3 that nonemergent, substance dualism is not utterly incoherent. What then would lead us to prefer panpsychism over dualism, but a commitment to naturalism? Certainly, speculating that the basic physical particles that compose everything are themselves conscious is no simpler or more parsimonious an explanation than substance dualism. One might also worry whether panpsychism is recognizably a form of naturalism; a universe suffused with consciousness is difficult to reconcile with the causally closed, physically determined network of physical entities that is supposed to compose the naturalist's universe.[30] We might then think that the panpsychist solution is not entirely helpful to the emergentist.

Fortunately for the emergentist, there is some significant room for maneuver here short of entertaining panpsychism: "Since [an emergent entity] is causally generated by and dependent on an un-

29. See ibid., 1–74; Chalmers, *The Conscious Mind;* and Nagel, *The View from Nowhere,* for contemporary philosophers who at least entertain panpsychism.

30. See Moreland, *The Recalcitrant Imago Dei,* 139.

derlying system, it is ipso facto part of that system. It does not have … an existence independent from it. The question of how it could get 'hooked up' with that particular system does not properly arise."[31] The point here is that if emergence is indeed a fact, then psychological states are homogenous with the neurological constituents from which they emerge. It may seem counterintuitive that they are homogeneous but in fact they are, and we know this because of the fact of their emergence. In other words, "this properties-can-reproduce-after-their-own-kind-alone objection lacks justification,"[32] so unless the proponent of (3) can give some reason in principle why ordinary physical entities cannot have as among their natural powers the propensity to cause emergent, psychological features, there is no decisive reason to accept (3), and there is certainly no reason to go to the excess of panpsychism to avoid its consequences. Of course, if we isolate the constituents of the nervous system, we would not expect them to be the sorts of things that have a natural potential to cause psychological states, but the very same thing could be said about hydrogen and oxygen and the powers of water. That is, when considering oxygen atoms completely in precision from their compositional role in H_2O, nobody would expect them to have the natural potential to contribute to the power to dissolve salt, but we aren't subsequently concerned that the power to dissolve and the features of hydrogen and oxygen atoms are so radically heterogeneous phenomena as to rule out the possibility that the latter accounts for the former.[33]

31. O'Connor, *Persons and Causes,* 119. Please note that (a) O'Connor makes this remark when actually addressing the issue of how an emergent property could have an effect on its physical base, though his remarks, as I argue in this paragraph, likely say as much to defend the notion that a physical base can cause a nonphysical emergent property to come to be (and presumably continue to exist); and O'Connor himself does not defend a quantitative version of emergence. As we shall see, his strategy is to deny premise (1). Moreland argues that O'Connor's position is actually a version of panpsychism; see *The Recalcitrant Imago Dei,* 137–42.

32. O'Connor, *Persons and Causes,* 121.

33. C. D. Broad makes this point in his classic defense of emergentism, *Mind and Its Place in Nature.* See the excerpt from this text, "Mechanism and Its Alternatives," in *The Philosophy of Mind: Classical and Contemporary Readings,* ed. David Chalmers, 106–15

These consideration do not amount to a decisive rejection of premise (3). Certainly, if emergentism is true, then some physical entities have a natural power or potential to cause psychological states, but the question is whether emergentism is true. I take it that Searle and O'Connor (albeit in very different ways) claim that emergence is the best explanation available that avoids the problems of both materialism and nonemergent dualism. If that is correct, one might then say we should go with emergence, however strange or unlikely it might seem in itself. Strawson's point, however, is that emergence is so implausible in its own right, given the *seeming* radical difference in kind between physical and psychological properties, that we simply cannot take it seriously as an explanation. The problem is exacerbated by the fact that the anti-materialist arguments that provide part of the case for emergentism go a good way toward supporting Strawson's point; that is, psychological phenomena do seem, at least, to be fundamentally different from physical phenomena. It would help the emergentist's case a great deal if he could articulate a metaphysical model that would offer some insight into how essentially nonpsychological matter could have a natural potency to cause psychological states, but that is not something that seems terribly likely within the confines of the currently accepted mechanistic philosophy of nature, which O'Connor at least is not inclined to question.[34] Of course, none of this amounts to a decisive refutation of emergence either, so it seems that we don't at this point in our discussion have fully satisfactory reasons for or against (3).

The case for premise (1). Premise (2) may be plausible (or at least some emergentists are likely to grant it), and the case for (3) is really an open question, but these are concerns only if we have some reason to accept (1). Why would anybody take this to be true? Why, one might ask, must we take satisfying (a) or

(New York: Oxford University Press, 2002). See also Meehl and Sellars, "The Concept of Emergence."

34. See *Persons and Causes,* 108–9.

(b) as a necessary condition for emergence; why can't there be some third option? The reason is that it seems that the only other option is to say that emergence is just a brute fact, that is, when *x* emerges from *y,* that's just what happens and there is no deeper explanation as to why it does. Our only workable models for truly explanatory emergence are structural and quantitative emergence; it is just hard to imagine how else emergence might go. Thus, it's reasonable to conclude that the only option is to fit emergence into either (a) or (b) in order to avoid appealing to a supposedly brute fact. What's so bad about brute emergence? Galen Strawson makes the following point: "It is built into the heart of the notion of emergence that emergence cannot be brute in the sense of there being absolutely no reason in the nature of things why the emerging thing is as it is."[35] That is, if the base phenomena do not explain why the supposed emergent phenomena are the case, then why would we conclude that we have bona fide cases of emergence in the first place? There can be closely associated phenomena without emergence, so then what separates those accidental associations from emergence, if not some deeply explanatory relation? The notion of coincidental emergence seems to stretch what we could possibly mean by "emergence": "If someone says he chooses to use the word 'emergence' in such a way that the notion of brute emergence is not incoherent, I will know that he is a member of the Humpty Dumpty army and be very careful of him."[36] If emergence can be a brute, inexplicable fact, then anything might emerge from anything else so long as they are closely associated. We might then just as easily say that something emerges from nothing, which makes emergence a sort of *deus ex machina* or a magic bullet that can be used to justify just about any claim no matter how strange.[37] Leaving aside

35. Strawson, *Real Materialism,* 65.

36. Ibid.

37. Ibid., 66. This is also a point made very well in the ancient world by Epicurus; see Epicurus, "Letter to Herodotus."

Strawson's well-placed, even if less than charitable, dismissal of brute emergence, it would be helpful if the defender of such a possibility could give a noncontroversial example of a case of brute emergence besides consciousness. The fact of the matter is that nobody has convincingly done so as of yet, which leads one to worry that appeals to brute emergence are all too convenient cases of special pleading.[38]

If the coming to be of psychological states is a case of structural emergence or quantitative emergence, there clearly wouldn't be a need to appeal to any brute facts; it is clear that plywood and screws have properties sufficient to bring about containment, H_2O molecules have properties sufficient to bring about liquidity, atoms have properties sufficient to bring about solidity, collections of players are sufficient to bring about teams, and so on. In each of these cases, emergence isn't brute, but explanatory, and we can understand as much just by understanding the properties and structure of the base phenomena. In fact, given the microstructure and the essential properties of the entities involved, we can see that the emergent feature is a necessary consequence of the base phenomenon. Thus, if brute emergence is really otiose and there is no other readily available option for emergent explanation in addition to structural and quantitative emergence, then we have good reason to accept (1). Given the plausibility of (1) and (2), we then have some good reason to doubt emergentism.[39] Maybe there is

38. J. P. Moreland makes this point in *The Recalcitrant Imago Dei,* 10, 35, 137–42.

39. Thomas Nagel makes a similar case in "Conceiving the Impossible in the Mind-Body Problem," *Philosophy* 73, no. 285 (1998): 337–52. Other helpful critiques of Searle's position can be found in Moreland, *The Recalcitrant Imago Dei,* 32–36, and "Hume and the Argument from Consciousness," in *In Defense of Natural Theology: A Post-Humean Assessment,* ed. James Sennett and Doug Groothius, 271–96 (Downers Grove, Ill.: InterVarsity Press, 2005). Plantinga's arguments against supervenience are also quite helpful; see his contribution to *Knowledge of God,* 51–66. See Michael Tooley's reply to Plantinga (*Knowledge of God,* 184–205) for a sophisticated defense of a version of emergentism. Note that Tooley calls his position a form of property dualism, a term Searle refuses to use to characterize his own position, though many commentators are perplexed in what sense biological naturalism is not a kind of property dualism. See John Searle, "Why I Am Not a Property Dualist," in *Philosophy in a New Century: Selected Essays,* 152–60 (Cambridge: Cambridge University Press, 2008).

some other option for explaining emergence, but the problem is that we have no idea as to what it would even be like to have such an explanation. Thus, emergence theories of the mind seem like an attempt to gain a free lunch or at least the issuing of a very unlikely promissory; there is absolutely nothing about the chemical processes occurring within neurons with which we are remotely acquainted that in any way explains the fact of consciousness. Of course ignorance of an explanation does not entail that there is no such explanation, but we should ask ourselves how likely such an explanation is given our general understanding of the world. The probability, as far as we can tell, seems to be exceedingly low.

Some philosophers give in to despair and advocate a naturalistic version of *mysterianism.* This is the view that, although we have a strong hunch that consciousness is caused by a biological mechanism, there can be no scientific explanation of it (which for the naturalist is to say there can be no explanation for consciousness at all). Colin McGinn argues that we can be aware of our psychological states and the brain, but we can never perceive the relationship between consciousness and the brain any more then we can look at our own eyeball directly.[40] The mind-brain relationship is the lens through which we see everything else, so we can't step outside of it in order to put it under our gaze. There just are some things, even in a naturalistic universe, that we cannot know because of our own cognitive limitations. Even if one grants McGinn's point here, one might still ask why it is one would persist in being a naturalist, if indeed there are phenomena for which there can be no naturalistic explanation. One could just as easily see a nonnaturalistic version of dualism as equally plausible.

Searle's way of avoiding (1) is not to embrace mysterianism about the mind, but to question whether brute facts are really so

40. See Colin McGinn "Can We Solve the Mind-Body Problem," *Mind* 98, no. 391 (1989): 349–66. I will refer to McGinn's article as it appears in *Philosophy of Mind: Classical Problems/Contemporary Issues,* ed. Brian Beakely and Peter Ludlow (Boston, Mass.: MIT Press, 2006).

bad. For instance, in replying to arguments questioning the plausibility of neurological accounts of qualia because of an apparent lack of necessity in any such relations, Searle claims: "On the contrary, nature is radically contingent. I am with Hume in thinking that the conviction that nature must necessarily be the way it is is an illusion. It is just that nature turned out one way and not another way. Nature is full of surprises. So, I do not regard [this sort of] objection as in any way conclusive against the possibility of a neurobiological explanation of consciousness."[41]

We do not have the luxury of digressing on Hume's skepticism about causation (though the exercise is highly recommended!), but suffice it to say that Searle has attempted to defend his version of emergentism by arguing that there is no causal necessity whatsoever in nature.[42] Searle grants that there is no necessity between neurophysiological systems and their supposed emergent features, but that is just, as Searle's remarks above imply, because there is no necessity anywhere in nature; for example, we *might* find H_2O from which liquidity doesn't emerge under the same conditions that it normally does.[43] Thus, psychological states are neither more nor less mysterious than any other emergent property (or any other cause and effect relation for that matter). We just don't have the habit of thinking of it in this way yet, but eventually we will get used to it. In short, Searle

41. Searle, *Mind: A Brief Introduction,* 103. Note that Searle's apparent sympathy for a Humean account of causality evidenced in these remarks is difficult to square with remarks he makes elsewhere in the same volume; see ibid., 136–44.

42. The most accessible presentation of this view is sec. 4 of Hume's *An Enquiry Concerning Human Understanding.* Note that Searle is not endorsing the whole of Hume's causal skepticism, but only seemingly the notion that there is no causal necessity in nature.

43. We need to distinguish the denial of causal necessity that Searle is appealing to from the claim that causation is opaque in the sense that we can offer no model of it that I defended in chapter 2. The Humean position is that we are never justified rationally in believing that A causes B. My claim is that we are often justified to believe that A causes B, but we cannot explain causal relationships in terms of any more fundamental relationship. Also note that I am not claiming that Searle is taking on the whole of Hume's skepticism about natural causation, though the remarks quoted above seem to indicate that he is willing to go a good way down that path.

attempts to avoid Strawson's complaint about the incoherence of emergent brute facts, by rejecting the very notion of a necessity in nature; ultimately all facts are brute. Losing the very idea of natural necessity is quite a high price to pay for a theory of mind. Do you really think it possible that under standard conditions large quantities of H_2O molecules could fail to be a liquid or that plywood and screws might not cause containment when properly arranged?[44] If accepting emergentism requires us to answer this question affirmatively, one is well within her rights to look elsewhere for answers in the philosophy of mind.

O'Connor has proposed a reply to (1) that comes at a much smaller price for our common sense intuitions about natural necessity:

> I thus agree that the relation of causation between one event and another … is a species of metaphysical necessity. Now my objectors make another, more dubious claim: once we come to have a scientific understanding of a certain type of phenomena—say gravitational attraction—we can in some sense simply see that the one even had to follow the other. All genuine causal necessity, they maintain, must ultimately be transparent to the inquirer. This connection of true necessity with the inconceivability of things being otherwise is of course an old one. Nonetheless, it is implausible. At any rate, contrary to the objector's contention, we certainly don't see, in a transparent fashion, the causal necessity involved in even the most well-understood natural phenomena.[45]

O'Connor's point is that there are two ways we can understand the rejection of brute emergence that lies behind our premise (1):

(i) If *x* causes *y* in circumstance *C,* then it is necessary that *x* causes *y* in *C.*

(ii) If *x* causes *y* in *C,* then it is inconceivable to us that *x* not cause *y* in *C.*

44. Strawson makes this point well: "This whole process is underwritten by the wild radical-empiricist-inspired metaphysical irresponsibilities of the twentieth century that still linger on (to put it mildly) today and have led many, via a gross misunderstanding of Hume, to think that there is nothing intrinsic to a cause in virtue of which it has the effect it does" (*Real Materialism,* 66).

45. O'Connor, *Persons and Causes,* 118.

O'Connor believes that there is no reason to doubt, and indeed he affirms, (i); that is, he believes that causal relations are necessary, so he grants that if psychological states emerge from a certain physical base under certain conditions, then they necessarily so emerge. In short, he denies a sort of Humean skepticism about natural causes. That, however, does not entail (ii), which is clearly false. The fact that something is necessary and that we even recognize it as necessary does not imply that it is inconceivable to us that it be otherwise. "P is necessary" and "Not P is inconceivable" are not equivalent; for example, we discussed in earlier chapters that "Water is necessarily H_2O" does not entail "It is inconceivable that water is not H_2O." O'Connor correctly denies that "genuine causal explanations of an event will leave its connection to its cause unveiled and transparent to the informed eye."[46] That is, all causal explanations, according to O'Connor, are brute, in the sense that we could conceive things to have gone otherwise, but that in no way requires that they be brute in the sense that they are radically contingent or that the cause doesn't actually bring about the effect. Thus, the fact that we cannot see transparently that certain neurophysiological structures cause psychological properties to emerge (we could conceive otherwise) in no way implies that there is no such causal relation; and indeed in this way the emergence of psychological states would be in no way different from many fundamental causal relations in nature.

Take O'Connor's example of gravitational attraction. I don't doubt that this relation is necessary, given the natures of the physical substances that exist in the actual world, and that is perfectly consistent with the fact that it is, in some sense, conceivable that the laws of gravitational attraction falter. Notice, however, that in this case physicists are constantly contributing to a better explanation, for example, the curvature of space as part of

46. Ibid.

the story about gravity, of what we might originally have taken as a case of brute necessity. It is reasonable to expect that someday (even if we aren't all that close now) we will have a satisfying explanation of why gravitational attraction must follow certain laws given the kind of physical substances that exist in our universe. Of course, while that explanation is still outstanding, we still recognize gravitational attraction as necessary. The point here is that even in the case of what appear to be the brute necessities of nature, we still find some hope of eventually finding an explanation. The problem for the emergentist is that there isn't any plausible suggestion as to what sort of discovery would move us beyond positing a brute necessity at the root of the emergence of psychological states. What about the chemical processes in the nervous system could possibly provide a satisfying explanation of psychological states? Sure, we don't know just yet what exactly grounds the necessity of gravitational attraction, but we have some inkling as to where to look, and that makes our claim to have discovered such a necessary causal relation much easier to justify. We have no such inkling in the case of the supposed emergence of psychological states, which cannot help but weaken our confidence in the fact that there is such a causal relation. Some of these worries might be alleviated considerably if we had a broader metaphysical theory that helped us to understand how emergence might work, but that doesn't seem to be a promising prospect so long as we operate within the confines of mechanism. So, once again, this defense of emergentism is neither without merit nor entirely compelling.

Suppose that the emergentist could (contrary to what we have seen above) undermine all three of the premises of the argument we have discussed above. There is still one remaining question about emergence that philosophers have frequently raised: Why did consciousness evolve in the first place? If we are going to have a fully biological explanation of consciousness, we need to understand not only how brains cause consciousness,

but also why brains capable of consciousness would evolve in the first place. Do we have a biological explanation as to why some animals are conscious? Such an explanation would require us to specify how consciousness is advantageous for reproduction or survival. Some philosophers, for example Nietzsche, have argued that consciousness can only be thought of as superfluous for any real-world survival,[47] but Searle recommends where we might go for an evolutionary explanation:

> We can say that in conscious perception the organism has representations caused by states of affairs in the world, and in the case of intentional actions, the organism causes states of affairs in the world by way of its conscious representations. If this hypothesis is correct, we can make a general claim about the selectional advantage of consciousness: Consciousness gives us much greater powers of discrimination than unconscious mechanisms would have.[48]

I take it that Searle has in mind here an explanation for sensations, primarily in their qualitative aspect. The idea is that an animal that tasted (qaule) a certain color (quale) of berries and then felt nauseous (quale) might later recall the feeling associated with those and thereby act (agency) to avoid such things in the future. Being able to have experiences and organize them into some kind of coherent catalogue would seem to have a selective advantage. We might then expect evolution by natural selection to produce organisms with such conscious capacities.

David Chalmers, who does not doubt evolution by natural selection in general, has doubts about this sort of explanation of the qualitative aspect of psychological states: "People put forward speculation—maybe the function of consciousness is planning or decision-making or integrating information or whatever. But then as soon as such a hypothesis is put forward the questions just gets raised 'Why couldn't that have been done without consciousness? Why couldn't you just have had these brain pro-

47. See Friedrich Nietzsche, *The Gay Science,* trans. Adrian Del Caro (Cambridge: Cambridge University Press, 2001), 211–13.

48. Searle, *The Rediscovery of the Mind,* 107

cesses which produced that conclusion with no subjective experience anywhere?'"[49] Chalmers's point is that a zombie organism that has the same functional capacity as its conscious twin would do well as long as it "outputs" the appropriate behaviors. Consider two organisms, *A* and *B,* such that *A* has qualitative, first-person experiences and *B* is a zombie version of *A* that lacks qualia entirely. *A* and *B* would be functionally equivalent; they would output the same behaviors given the same inputs. The qualia or first-person perspective that *A* possesses, but that *B* lacks, adds nothing to the adaptive advantage of a certain mode of behavior, so there is no reason to suppose that *A* would survive or reproduce any better than *B.* Whether something experiences what it is like to run away from a saber-toothed tiger while doing so in no way makes a difference for whether it is eaten by the predator. All that matters is the *running,* not the *feeling of running.* So what good does consciousness really do an organism? Assuming this sort of zombie scenario is a legitimate possibility, qualitative consciousness (sensation) does not do any work that could not be done by functionally equivalent nonconscious features of the organism, so it is not something we would expect to be preserved by the process of natural selection. The point here is not that other features might have been just as adaptively advantageous as psychological states. Indeed, there might have been other equally or even more advantageous capabilities for birds than their power of flight, but that of course does not imply that we cannot explain the power of flight in terms of its adaptive advantage.[50] Rather, the point is that consciousness doesn't seem to do any adaptive good at all; whether or not an animal feels a certain way makes no difference as long as it functions in a way that is adaptively advantageous, so it is unclear that the qualitative aspects of consciousness really make any contribution to survival. Maybe

49. David Chalmers, "I'm a Conscious Zombie," in Susan Blackmore, *Conversations on Consciousness: What the Best Minds Believe about the Brain, Free Will, and What It Means to Be a Human* (New York: Oxford University Press, 2006), 47.

50. Thanks to Patrick Toner for offering me this example.

consciousness is a useless feature that "piggybacks" as a coincidental side effect of some nonconscious, but adaptively advantageous, feature of certain organisms. Even if that were the case, we still would not have an explanation as to why consciousness arose in the first place, even if it accompanies some other useful property. What about the useful (nonconscious) property causes consciousness? Moreover, nobody has any idea on what nonconscious feature of the central nervous system consciousness is supposed to piggyback. Indeed, we haven't any idea where to look for such a feature, because we are aware of no physical feature of organisms that is necessarily connected to consciousness. Thus, the introduction of consciousness still remains mysterious, and this is a mystery that needs to be addressed before we can accept naturalistic versions of emergentism.[51]

None of these criticisms of emergentism is decisive, but they do collectively serve to undermine much of the case in its favor. If emergentism were the only game in town, then one might say it is acceptable, but a case could be made that nonemergent dualism is, though not without its difficulties, equally acceptable. So why would somebody go along with emergentism? Searle answers this question when he says that any form of supernaturalism

> is not an option. It is not simply up for grabs along with a lot of competing worldviews. When we encounter people who claim to believe such things, we may envy them the comfort and security they claim to derive from these beliefs, but at bottom we remain convinced that either they have not heard the news or they are in the grip of faith. We remain convinced that they must separate their minds into compartments to believe such things. Given what I know about how the world works, I could not regard their views as serious candidates for truth.[52]

Searle makes the point I have alluded to throughout this section: naturalism is not a conclusion that philosophers come to based on what they discover in the philosophy of mind, but a premise implicit in many of their arguments. That is, there is nothing

51. Thanks to W. Matthews Grant for raising these issues.
52. Searle, *The Rediscovery of the Mind*, 91.

in the philosophy of mind that truly demands naturalism, and actually much to recommend against it, but if one has a pre-philosophical (at least pre-philosophy of mind) commitment to naturalism, then she is going to be willing to do conceptual gymnastics to account for mind in naturalist-friendly terms. McGinn puts it succinctly: "Resolutely shunning the supernatural, I think it is undeniable that it must be in virtue of *some* natural property of the brain that organisms are conscious. There just *has* to be some explanation for how brains subserve minds."[53] Why think, even contrary to what seems to be evident in the philosophy of mind, that there must be such a *naturalistic* explanation?

Thomas Nagel's reply to this question is frank, as he admits to a "cosmic authority problem" that stems from a fear of religion. Nagle confides "I hope there is no God! I don't want there to be a God; I don't want the universe to be like that," and he conjectures that "this cosmic authority problem is not a rare condition and that it is responsible for much of the scientism and reductionism of our time."[54] I will not address Nagel's psychological explanation of other philosophers' commitment to various naturalistic explanations of mind, and my point is not to argue that the failure of naturalistic theories of mind strongly supports theism.[55] Rather, the point I want to emphasize is that broader philosophical commitments seem to play a role in the philosophies of mind defended by some naturalists. Nagel, at least, seems to admit that his commitment (maybe even a quasi-religious commitment) to naturalism determines his positions in some areas of philosophy. Someone would not entertain many of the views we have discussed in this chapter unless she had a prior commitment to naturalism. However, if the best the naturalist can do is

53. McGinn, "Can We Solve the Mind-Body Problem," 323.

54. Thomas Nagel, *The Last Word* (New York: Oxford University Press, 2001), 130.

55. I am not, however, denying that a failure of naturalism about persons is evidence in favor of theism. Moreland makes such a case succinctly in *The Recalcitrant Imago Dei,* 1–15. C. S. Lewis likewise makes an argument for theism based on the failure of naturalistic explanation for the mind in *Miracles,* 37–51.

some version of a mysterianism in the philosophy of mind, and it is only because of these broader philosophical, or even religious, commitments that one would persist in being a naturalist, then we do well to call that commitment into question.[56] What more would need to be done to defeat naturalism than to show that naturalistic explanations of the mind are deeply dubious while pointing out that there are plausible nonnaturalist alternatives?

Emergentism and Abstract Thought

Our results in the previous section were inconclusive: we arrived at neither decisive refutation nor overwhelming strong support for emergentism. In this section, I want to outline a stronger case against a broad application of emergentism: *it is impossible for certain thoughts to emerge from* (*or be causally reducible to*) *any physical organ.* This argument will leave the issue of the emergence of sensations an open question, but it purports to show that the emergence of thoughts is metaphysically impossible. I call this line of reasoning the "Thomistic argument," not because it can be found verbatim in the writings of St. Thomas Aquinas, though he does say some things along these lines, but because it is an argument that appears in various forms in the writings of contemporary Thomistic philosophers, followers of St. Thomas.[57] Let's begin the presentation of this argument with the following example:

56. O'Connor is an emergentist, but he is certainly no naturalist; he actually vigorously defends theism. See O'Connor, *Theism and Ultimate Explanation.* O'Connor does, however, take it that every "*macro*-level phenomenon has arisen through entirely natural microphysical causal processes and that its existence continues to causally depend on processes of this kind" (*Persons and Causes,* 94), as an unquestioned first principle. We might wonder, however, why we should accept this principle, given the success, which O'Connor grants, of anti-materialist arguments in the philosophy of mind.

57. For St. Thomas's various arguments against materialism, see *Summa contra Gentiles,* trans. James Anderson (Notre Dame, Ind.: Notre Dame University Press, 1956), book 2, ch. 50; *Summa theologica,* Pt. I, Q. 75., A. 2, trans. Fathers of the English Dominican Province (New York: Benziger Bros., 1947), and *Commentary on Aristotle's De Anima,* trans. Kenelm Foster, O.P., and Sylvester Humphries (Notre Dame, Ind.: Dumb Ox Books, 1954), lec. 7. For recent Thomistic arguments see Haldane, "A Return to Form in the Philosophy of Mind," "The Metaphysics of Intellect(ion)," *Proceedings of the Ameri-*

(5) Patrick thinks that *the robin is red.*

In (5), Patrick thinks that a certain *individual* (the robin) instantiates a *universal* (red). A universal is a type or kind, a repeatable entity (something that can be had by multiple individuals simultaneously). For example, the robin and the fire hydrant both have *red* simultaneously. The *red robin* and the *red fire hydrant* are particulars, which unlike universals cannot be instantiated or had by other things. The following propositions also refer to thoughts involving universals:

(6) Jack thinks that the thing in the tree is a bird

and

(7) Martha thinks that the cookie is tasty.

Both (6) and (7) claim that some particular is an instance of a universal (*bird* and *tasty* respectively). In fact, the vast majority of our thoughts involve attributing universal characteristics to individuals. In short, thoughts (or at least some of them) have universals as part of their content.

So far this seems pretty innocuous for the emergentist, but when we stop to wonder about what universals must actually be like, things become a bit more complicated. Russell Pannier and

can Catholic Philosophical Association 80 (2006): 39–55, and *Reasonable Faith* (New York: Routledge, 2010); Eleonore Stump, *Aquinas* (New York: Routledge, 2003), 200–207; James Felt, SJ, *On Human Knowing: A Prelude to Metaphysics* (South Bend, Ind.: University of Notre Dame Press, 2005), 61–77; George and Lee, *Body-Self Dualism in Contemporary Ethics and Politics,* 52–68; David Braine, *The Human Person: Animal and Spirit* (Notre Dame, Ind.: University of Notre Dame Press, 1992), 447–79; David Ruel Foster, "Aquinas on the Immateriality of the Intellect," *Thomist* 55 (1991): 415–38; Russell Pannier and Thomas D. Sullivan, "Consciousness and Intentional Awareness of Instantiables," in *Mind and Its Place in the World: Non-reductionist Approaches in the Ontology of Consciousness,* ed. Alexander Batthyany and Avshalom Elitzur, 77–100 (Frankfurt: Ontos Verlag, 2006); James Ross, *Thought and World: The Hidden Necessities* (Notre Dame, Ind.: University of Notre Dame Press, 2008); Ric Machuga, *In Defense of the Soul: What It Means to Be Human* (Grand Rapids, Mich.: Brazos Press, 2002); Mortimer Adler, *Intellect: Mind over Matter* (New York: Macmillan Publishing, 1990); Gyula Klima, "Aquinas on the Materiality of the Human Soul and the Immateriality of the Human Intellect," *Philosophical Investigations* 32, no. 2 (April 2009): 163–82. My presentation of the Thomistic argument has been influenced by many of these efforts, though I do not follow any of them in particular in what follows in this chapter. I make what could be broadly called a Thomistic argument against emergentism in "Realism, Nominalism, and Biological Naturalism."

Thomas Sullivan use the example of the thought "this is a neuron" to make the relevant point about universals:

> In grasping the complex thought, you grasp the constituent concept *neuron.* The characteristic neuron is an instantiable, i.e. there can be more than one instance of a neuron. The important point for our argument is that unlike particular instances of neurons, e.g., the neurons in your head that are firing now, the instantiable characteristic neuron has no spatial location. It is pointless to ask "Where is *neuron*?" The same is true of other instantiable characteristics such as *validity* or *consternation.* No one asks "How many meters separate validity from consternation?" or "How wide is equanimity?"[58]

Pannier and Sullivan seem to take "concept" in the same sense we are assigning to "universal," and their point is that the universal *equality,* as opposed to two equal particulars, is nonphysical.[59] *Equality* itself doesn't exist *any place,* whereas all physical objects are spatially located (even if not extended). It makes sense to ask, "Where are the two equal sticks?" but it is nonsense to ask, "Where is *equality*?" Likewise the universal *neuron* is unlike any particular neuron, because it is nonphysical.

This insight has little of interest to say about the possibility of an emergentist account of sensations. There is no reason to say that one must, for example, understand the universal *pain* in order to have *a pain;* squirrels no doubt have pains, but there is no reason to think that they understand the relevant universal. Moreover, one may be in a qualitative state of seeing blue, without grasping the universal blue. For example, babies presumably have the qualitative feel of *seeing blue,* before they know what *blueness* is. Thoughts, however, involve universals, and as such they give us very good reason to doubt they can be given an emergentist explanation.[60] When I think, to use Sullivan and Pannier's example:

58. Russell Pannier and T. D. Sullivan, *Great Thinkers on Great Questions,* ed. Roy Abraham Varghese (Oxford: One World Publishing, 1998), 137.

59. This is a classic philosophical example Plato made famous in his dialogue *Phaedo.*

60. See John Frederick Peifer, *The Concept in Thomism* (New York: Bookman, 1952),

(8) This is a neuron,

I am in a state that involves the *universal neuron.* Thus, the thought expressed in (8) cannot emerge from any *particular* neurophysiological event no matter how complicated, because it involves a *universal.* All physical objects are particular, and thoughts have universal content. John Haldane makes the crucial point motivating this argument when he points out that thought "is structured by universals and universals exist as such apart from (empirical) matter."[61] Thoughts involving universals cannot satisfy either of the conditions for emergence we considered in the last section. On the one hand, heap all of the structural complexity you may want into the neurophysiological set-up for our thoughts, but you will never get something universal; no addition of particulars, however structured, ever amounts to a universal. On the other hand, we cannot appeal to a proto-universality in the fundamental nature of physical particles because physical particles are particulars, and there is no such thing as being *a little bit universal.*

We might state the argument in a simple form as follows:

(9) If thought emerges from a neurophysiological process, then there should be no feature of thought that cannot in principle be accounted for by features of that neurophysiological process. (premise)

(10) Thought has universal content. (premise)

(11) Because they are particulars, neurophysiological process have no features that can account for universal content. (premise)

Therefore:

(12) There is a feature of thought that cannot be accounted for by a neurophysiological process. (from 10 and 11)

Therefore:

(13) Thought does not emerge from a neurophysiological process. (from 9 and 12)

113–31 for a detailed discussion of the distinction between sensation and thought in Thomism, wherein the latter but not the former requires universal concepts. Peifer's book is highly recommended to anybody interested in an in-depth study of St. Thomas's epistemology and philosophy of mind.

61. Haldane, "A Return to Form in the Philosophy of Mind," 58.

Premises (12) and (13) follow validly from other premises, (9) is noncontroversial, and I gave reasons for (11) in the last paragraph. There is a very important objection to (10). This premise presumes a partisan position in one of the history of philosophy's most abiding controversies. There are philosophers, called *nominalists,* who doubt that there are universals, and they doubt that we need the notion of universal content to makes sense of our thoughts. According to the nominalist, all that exists is particular and nothing is really multiply instantiable. "Nominalism" is derived from the Latin for "name." Names, for example, "Jim Madden," are phrases that simply refer to some object without describing it as having some characteristic. For example, "Jim Madden" does not tell you anything about me; it's only a way of pointing to me or getting my attention. Nominalists, at least in the most extreme form, believe that all of our words are really names. According to a strict version of nominalism, the red robin and the red fire hydrant are thought to have nothing strictly in common but that we use the word "red" to refer to them. There is no shared universal characteristic, *red,* and the word "red" is no different from the proper name "Jim Madden." We cannot even begin to scratch the surface of the debate between nominalists and *realists* (philosophers who accept the reality of universals), but suffice it to say that if you think that the two trees outside the window of my office have some intrinsic characteristic in common in virtue of which they are trees, then you are likely some kind of realist, and therefore you likewise have reason to accept (10). If you don't think the two individual trees have *treeness* in common, then you are nominalist, and you have good reason to doubt (10).[62] I myself am a realist, and I would

62. Introductions to the realism-nominalism debate can be found in H. H. Price "Universals and Family Resemblances," in *Classics in Analytical Metaphysics,* ed. Larry Lee Blackman (Lanham, Md.: University Press of America, 1984); Armstrong, *Universals;* and Loux, *Metaphysics.* Peter Kreeft gives many accessible arguments against nominalism in *Socratic Logic* (Notre Dame, Ind.: St. Augustine's Press, 2005), 35–47. For nominalist replies to the Thomistic argument in particular see Wilfrid Sellars, "Being and Being Known," in *Science, Perception, and Reality,* 41–59 (Atascadero, Calif.: Ridgeview, 1991), and Hilary Putnam, *Words and Life* (Cambridge, Mass.: Harvard University Press, 1994).

ask any reader who seriously entertains nominalism to consider whether she really believes that the trees really have nothing in common or that our categorizations of things are merely arbitrary acts of naming. I must admit that I find that hard to believe.[63]

Some philosophers might worry that there is an ambiguity regarding what is meant by "universal content" in (10) and (11) as I have stated them.[64] On the one hand, one could mean that thoughts have universal content in the sense that they have universals as components or constituents; that is, universals are in some sense *parts* of thoughts, such that thoughts are themselves, in part, universals. On the other hand, one could mean that thoughts have universal content in the sense only that they refer to universals, though they do not have universals as components.[65] Although I favor the former option (because it more readily squares with a

John Haldane replies to Putnam in "On Coming Home to (Metaphysical) Realism," *Philosophy* 71 (1996): 287–96. I argue in "Realism, Nominalism and Biological Naturalism" that one need not actually defend realism in order to defeat Searle's biological naturalism, only that either realism or nominalism must be true. That argument is too complicated for this context, so I refer the reader to "Realism, Nominalism, and Biological Naturalism."

63. Often the primary objection to realism is that it forces us to believe in a great many abstract entities, and it forces on us the arduous metaphysical task of articulating the relation through which individuals participate or exemplify universals. Extreme forms of realism take universals to exist independently of any particulars at all, even minds thinking about them. Some *extreme realists* believe that universals exist in possible worlds in which no individuals instantiate them, and indeed on this view there are infinitely many uninstantied universals existing now: one for each property things might have, all of which are somehow related to any concrete particular that actually exemplifies them. The nominalist's impatience with extreme realism is understandable; if we could get by without so many abstract objects in our metaphysics, we would do well to do so. There are, however, *moderate realists* who do not believe that universals have real existence outside of thoughts, and it is this sort of realism that St. Thomas actually defended. A moderate realist, like St. Thomas, is not committed to a universe crowded with unexemplified universals, so long as there are intellects capable of having universals as objects of thought. Moderate realism has metaphysical consequences (thought is not a physically emergent process), but it avoids the traditional nominalist objection of ontological overpopulation. See Joseph Owens, "Common Nature: A Point of Comparison between Thomistic and Scotistic Metaphysics," *Mediaeval Studies* 19 (1957): 1–40, for a very helpful discussion of St. Thomas's moderate realism. Peifer, *The Concept in Thomism,* 132–89, is also very helpful in clarifying the Thomistic doctrine regarding universals. Roderick Chisholm serves as a good example of a contemporary version of extreme realism in *Person and Object,* ch. 4.

64. Thanks to W. Matthews Grant for raising this insightful objection.

65. I take it that this distinction is roughly the same as the distinction Robert Pasnau makes between something being in the intellect *concretely* and being in the intellect *inten-*

moderate version of realism wherein universals exist primarily *in* the intellect),[66] I believe that on either interpretation (10) and (11) together entail (12). Certainly, if universals are components of thoughts, then thoughts cannot emerge from processes involving particular entities, as I argue in the previous paragraph. If we take it that universal content is just a matter of referring to a universal, then we would need to spell out what it means to do so. It is possible to refer to a particular concrete object without a representation or description of such an object, because one could fix the reference literally by pointing to it; or if someone is not directly acquainted with it, then his relation to others who are so acquainted could fix the reference with no need for a description or representation. This will not do for universals, because they are abstract, and therefore one cannot have acquaintance with them without representation or description. Acquaintance with an abstract object certainly involves understanding or conceiving it, and those are representational notions. Whatever counts as a description or representation of a universal, it cannot be a matter of the proper arrangement of physical constituents, since an arrangement of particulars could never represent it *as universal.* In short, a neurophysiological process cannot be the basis for a state that refers to a universal. Thus, even the weaker versions of (10) and (11) entail that thoughts cannot emerge from neurophysiological processes, because of their universal content.[67]

---

tionally. I also take it that the following remarks are sufficient to answer his criticism of the Thomistic argument from abstract thought. See Robert Pasnau, *Thomas Aquinas on Human Nature: A Philosophical Study of Summa Theologiae 1a 75–89* (New York: Cambridge University Press, 2002), ch. 2. John Haldane offers a reply to Pasnau's criticism distinct from the line I follow below in "The Metaphysics of Intellect(ion)," 51–54.

66. See note 50.

67. We will discuss this argument in some detail again when we consider the Thomistic understanding of human souls in chapter 8.

Emergentism and Agency

Finally, as I closed the last chapter with concerns over materialism and agency, I conclude our discussion in this chapter by raising similar concerns regarding emergentism. Even though psychological states are, according to the emergentist, real and capable of exerting a downward causal influence on the neurophysiological processes from which they are supposed to emerge, we would expect they are nevertheless physically determined by what occurs at the neurophysiological level. The problem is that the will to pursue a certain good is a system-level feature of what occurs at the basic, neurophysiological level, which is itself entirely causally determined. Remember all psychological states (and psychological agents) are supposedly causally reducible, to use Searle's term, to their neurophysiological foundations, and therefore psychological states have no causal powers that they exercise independently of the neurophysiological system from which they are supposed to emerge. As Searle puts it: "the causal powers of consciousness and the causal powers of its neuronal base are exactly the same."[68] It is far from clear how we can say that our experience of ourselves as moral agents is anything but illusory when Searle supposes that all of our conscious states are exhaustively determined by processes run by the laws of physics.

Searle is well aware of this problem, and he entertains two options for the emergentist. She might just bite the bullet and argue that our experience of being free, responsible moral agents is indeed illusory; our consciousness of agency is, in the end, epiphenomenal. Searle thinks that this option is indeed quite unacceptable for two reasons: (a) We can't literally believe that we are not agents. He gives here the example of deliberating over attractive menu items in a restaurant; one can't just plead determinism and wait to see what she orders for dinner, because the "the refusal to

68. Searle, *Mind: A Brief Introduction,* 89.

exercise free will is only intelligible to you as a refusal if you take it as an exercise of free will."[69] Intentional, and seemingly free, action is something we inevitably engage in, so it seems that the implicit belief that we are agents is unavoidable to us. (b) Searle claims that epiphenomenalism about agency "seems to me to run against everything we know about evolution. It would have the consequence that the incredibly elaborate, complex, sensitive, and—above all –biologically expensive system of human and animal conscious rational decision making would actually make no difference whatever to the life and survival of the organisms."[70] In other words, it would be strange indeed for evolution to go to such lengths to produce a highly complicated and demanding system (it takes a lot of calories to sustain the appearance of free agency!) that does nothing more than give us the illusion that we are free moral agents, even though we really have no such power.

The other option for the naturalist is to appeal to fundamental quantum indeterminacies in the brain that allow for truly random events, which make it such that some of our thoughts are *completely undetermined.* Thus, William's decision to order the macaroni and cheese instead of the hotdog might be free, even though it is determined by events in the brain, because those events are themselves the products of truly random events at the most basic physical level. Searle doubts this hypothesis because we have no evidence that quantum indeterminacy is reflected at a macro-level, and "even assuming we had a quantum mechanical explanation of consciousness, how do we get from indeterminism to rationality? If quantum indeterminacy amounts to randomness then quantum indeterminacy by itself seems useless in explaining the problem of free will because free actions are not random."[71] In other words, random events do not make for freedom; such events would be beyond our control, and cer-

69. See Searle, *Freedom and Neurobiology,* 43.
70. John Searle, *Rationality in Action* (Boston, Mass.: MIT Press, 2003), 286.
71. Searle, *Freedom and Neurobiology,* 75.

tainly any sound account of free agency will allow us to assign responsibility for our actions to ourselves. Free actions are not determined by prior physical events, but they are *explained* by the good the agent pursues, so an appeal to randomness of some physical events is far from delivering moral agency. Searle himself is skeptical as to whether we can ever recover the notion of moral agency if we assume biological naturalism, but he himself does not think we can sensibly give up on the idea that we are bona fide moral agents. Thus, like the reductive materialist, the biological naturalist faces a serious dilemma between maintaining her version of naturalism and the common human intuitions about our status as moral agents. It is frankly hard to believe that anyone can seriously entertain giving up the latter.

O'Connor believes that there is a fairly simple solution to Searle's dilemma available to the emergentist: "Given that there is nothing inconsistent about the emergence of an 'ordinary' causal property, able to causally influence the environments in which it is instantiated, it is hard to see just why there could not be a variety of emergent property whose novelty consists in enabling its possessor to directly effect changes at will (within a narrowly limited range and in appropriate circumstances)."[72] In other words, if indeed there are bona fide emergent properties or causal powers (though I don't believe O'Connor has really established as much), there doesn't seem to be anything that would bar the possibility that among them is the power to determine the will to pursue certain goods independently. O'Connor's point is that if we take emergence seriously, Searle's concern doesn't even arise. Emergent powers have their physical bases as necessary conditions for their existence and exercise, but that doesn't entail that they are completely determined. Indeed, if we allow that emergent powers are truly novel, then nothing bars us from concluding that they are not subject to the same deterministic liabilities as the physical

72. O'Connor, *Persons and Causes,* 121.

structures from which they emerge. One may worry whether truly nonphysical substances or properties can actually interact with the physical world, but we have seen plausible strategies for how a dualist (certainly including the emergent property or substance dualist) can effectively assuages these worries. In short, given the assumption of O'Connor's version of emergence, the notion of moral agency is plausible.[73]

There is, however, a deeper problem for emergentist accounts of rational agency looming here. Return to the example we discussed at the end of chapter 5 regarding Jack learning the logical law, *modus ponens.* As we found in that discussion, if Jack is going to be a bona fide rational agent in his application of *modus ponens,* he cannot do so merely as a consequence of prior brainwashing, habitual associations of ideas, parroting of patterns of linguistic behavior, or some such. Rather, he needs to understand *modus ponens* and be able to apply it in diverse cases; otherwise he will be no better than the man in the Chinese Room with respect to logical operations; he might go through the motions of reasoning, but he does not actually have this particular mode of rationality. However, grasping *modus ponens* requires that one can abstract the universal law from all of its particular instances. For example, consider the two following arguments:

(14) If something is a horse, then it is a mammal.

(15) Brownie is a horse.

Therefore:

(16) Brownie is a mammal.

And:

(17) If something is a tortoise, then it is a cold-blooded.

(18) Natalia is a tortoise.

73. This of course assumes either that rational agency and moral agency can be separated or that the emergentist can account for the former as well as the latter. We will see in the following paragraphs and the final chapter of this book that there are good reasons to doubt both of these assumptions.

Therefore:

(19) Natalia is a cold-blooded.

If Brendan recognizes (14)–(16) as deductively valid, but does not recognize (17)–(19) as valid, then he does not understand *modus ponens* (assuming that he knows what all the words mean). The point is that logical understanding, which is a necessary condition for rational agency, requires that one has a grasp of logical relations or rules in abstraction from the particulars to which they apply. That is, somebody who understands and actually reasons by *modus ponens* must grasp (at least implicitly) that both of the above arguments are instances of the logical law:

If P, then Q.

P.

Therefore:

Q.

Logical laws, such as *modus ponens,* are universals, in the sense that we have discussed above. We saw above in our discussion of the Thomistic argument that the emergence of universal content entirely from the constituents of a physical system is deeply implausible, if not simply impossible. Emergentism is thus unable to account for rational agency, because it cannot give an account of the universal content of logical reasoning.[74]

An emergentist might reply here that we might appeal to evolution by natural selection to save the day. That is, wouldn't organisms that are capable of bona fide logical reasoning have an adaptive advantage over organisms that have corresponding false beliefs? Surely, getting things correct is some advantage to survival! Supposing that it is correct that rational agency is an adaptive advantage, it is far from clear how evolution by natural selection can help address the root problem for the biological

74. Presumably these considerations would likewise undermine emergentist attempts to account for moral agency inasmuch as freely pursuing a good entails an *understanding* of what that good is.

naturalist. The issue is that it seems impossible to get universal content from the particular constituents even of an immensely complex physical system like the human brain. Adding the fact that such a system was developed over the span of eons of time through incremental changes, contributing to the fitness of the organism does nothing to bridge the metaphysical gap between the particular and the universal. In short, even in light of its many theoretical virtues, evolutionary biology cannot shed any light on how universal content might emerge from a system of particulars.

Leaving aside the metaphysical case against the emergence, some philosophers wonder whether there is any reason to believe that evolution by natural selection would be at all likely to produce rational agents. This issue is the source of a great deal of controversy, ever since Alvin Plantinga presented an argument to the effect that naturalism cannot help but fall into a general sort of skepticism.[75] Plantinga argues that if we assume that naturalism is true, then it is no more likely than not that we would have a belief-forming mechanism aimed at true beliefs. Why? Suppose you have two cavemen, Will and Jack, in a burning forest. Will believes "The molecules composing my body are about to undergo a sudden increase in kinetic energy which will cause then to dissipate quickly, thereby bringing about my painful demise," and as a result he leaves the forest as soon as possible. Jack, on the other hand, believes "If I stay here the evil fire monkeys will descend from the trees and drag me off to their dungeons for all eternity," and as a result he leaves the forest as soon as possible. Notice that either way, the cavemen vacate the burning forest and therefore survive to reproduce. Moreover, note that for

75. Plantinga's most recent and accessible presentation of this argument is in *Where the Conflict Really Lies*. See also Plantinga and Tooley, *Knowledge of God* and Plantinga and Daniel Dennett, *Science and Religion: Are They Compatible?* (New York: Oxford University Press, 2011), for discussions of this argument in a debate format. For more criticisms and Plantinga's replies, see James K. Beilby, ed., *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism* (Ithaca, N.Y.: Cornell University Press, 2002).

every true belief about the forest fire we could construct dozens of false but effective beliefs that would lead a caveman to take life-saving actions. On the assumption of a naturalistic version of emergentism, Plantinga claims that it is not likely that we are intellectual agents capable of forming true beliefs for good reasons, which means that biological naturalism is subject to a self-defeating skepticism no less than functionalism. Plantinga's argument is the center of fierce contemporary controversy, and even some thinkers sympathetic to nonnaturalist approaches are quite critical of his position.[76] I raise the argument only to point out that even if the emergence problem could be addressed, it nevertheless is an open question, at least to many contemporary philosophers, as to whether naturalistic emergentism could indeed account for rational agency.

76. For example, see Bernardo Cantens, "Cognitive Faculties and Evolutionary Naturalism," *Proceedings of the American Catholic Philosophical Association* 80 (2006): 201–8; and Michael Bergman, "Commonsense Naturalism," in Beilby, *Naturalism Defeated?* 61–90.

CHAPTER 7

BEFORE THE PHILOSOPHY OF MIND— THE PHILOSOPHY OF NATURE

Our discussion thus far might well amount to grounds for philosophical despair. The most promising versions of materialism cannot account for the reality of the qualitative aspect of our sensations, the intentional aspect of our thoughts, or the fact that we are intellectual and moral agents. The prospect of a naturalist, emergent property or substance dualism is not entirely without traction, but it is difficult to square with the broader metaphysical views held by most contemporary philosophers of mind, and there are strong arguments purporting to show that the universal aspect of thought and subsequently intellectual and moral agency cannot in principle be accounted for in terms of emergence.

At the end of chapter 3 I argued that, even though it is plagued by difficulties, nonemergent dualism is *believable*. What I mean here is that there is no "knock-down" refutation of dual-

ism, such that it would be irrational or even just unreasonable to defend it. Certainly, issues regarding mind-body interaction are not perfectly clear, but we discussed that there are some promising strategies for the dualist to answer these worries. At the very least, the mind-body interaction problem is not so difficult as to undermine otherwise compelling arguments for dualism. We also noted that the nonemergent dualist has difficulties when it comes to explaining the obvious psycho-physical dependence between consciousness and the brain, the status of nonhuman animals as bearers of psychological states, and the morally significant experience we have of ourselves as living human bodies, as opposed to minds or souls contingently associated with human bodies. None of these concerns, however, is an absolute "dealbreaker" for the dualist, and there are interesting ways of treating these problems proposed by contemporary dualists. Of course, the nonemergent dualist will need to tell increasingly tedious stories to account for a significant range of facts contrary to what we would most likely expect, should nonemergent dualism be true, but there is nothing incoherent about any of these stories. In short, even though nonemergent dualism raises philosophical perplexities, we should see it as a live option in the philosophy of mind.

If we contrast nonemergent dualism with naturalism in its various forms we have discussed so far, the former comes out favorably at the expense of the latter. Dualism might require us to indulge in some tedious just-so stories, for example, "when I'm drunk my mind functions just fine, it's just that my brain communicates bad information," but it is by no means self-defeating. As we have seen in chapter 5, it is very difficult to sustain any version of materialism without "leaving something out," that is, qualia, intentionality, or agency, which is at least paradoxical, because it seems that our engagement in the philosophy of mind, even if in doing so we are defending materialism, is itself among the phenomena that materialists seem to leave behind. We saw in chap-

ter 6 that similar paradoxical results still plague the naturalist versions of emergence because of the impossibility of giving an account of the universal aspect of thought. Since nonemergent dualism, despite its many problems, is free of these paradoxes, given the choice between only nonemergent dualism and any version of naturalism we have discussed (and my hope is that we have discussed all of the most important versions of naturalism), nonemergent dualism is the more satisfactory philosophy of mind.

You may very well be disappointed to find that we have done a great deal of difficult philosophical work in the preceding pages without much to show for the effort. If the philosophy of mind at best can recommend to us a position wrought with difficulties of its own solely on the basis of some highly contentious arguments offered in its favor and the fact that its competitor theories lead to intractable paradoxes, it would seem that we have grounds for intellectual despair. One might conclude that we would have done better just to leave well enough alone, rather than considering the litany of unsatisfactory theories of mind we have discussed in the preceding pages of this book. I agree that there is little hope for a fully satisfactory *philosophy of mind:* given the way the debate about the mind-body problem is typically framed, it is unlikely that we can ever avoid the dead ends we have encountered. This conclusion, however, does not rule out the possibility of a broader philosophical perspective that leaves behind the problems that give us pause when considering the standard positions in the philosophy. In fact, I plan to outline such a position in this and the following chapter.

You will remember, as we discussed in chapter 2, that both nonemergent dualism and materialism share a common assumption, mechanism; *all of the properties of physical objects* (*living and nonliving, conscious and nonconscious, etc.*)*, can be accounted for in terms of mass, motion, charge, and so on, and therefore the laws governing these properties can give a complete explanation of all physical*

occurrences that can be explained. In other words, the whole story about physical objects, and the natural world in general for that matter, can in principle be told entirely in terms of physical properties and the laws of nature. As we have seen, the proponent of mechanism must accept either dualism, by granting that there are in fact psychological states, though they occupy a nonphysical realm, or some form of materialism, by either denying that there really are psychological states at all or claiming that psychological states are identical to or supervenient on physical entities. Emergentists and the panpsychists question mechanism, or at least would revise it beyond recognition. Similar to nonemergent dualists, they both affirm the reality of psychological states while doubting that they are identical to or supervenient on what are typically thought to be paradigmatic physical properties. The emergentists and panpsychists depart from the standard mechanist position when they claim that irreducibly nonphysical, psychological states (or even psychological substances according to certain emergentists) are fully fledged aspects of the natural world; in the case of the emergentist this is because psychological states emerge from the properly organized physical constituents, whereas for the panpsychist psychological properties are among the basic sui generis properties of matter. The problem with these revisionary projects is that they give the appearance of ad hoc gerrymandering in the service of preserving a naturalist account of psychological states come what may. If, for example, the emergent nonphysical phenomena were what we would expect as a consequence of a broadly nonmechanistic philosophy of nature that we hold for reasons independent of the philosophy of mind, then this appearance of special pleading would be erased.

Mechanism is certainly the dominant philosophy of nature in the modern world, but it holds unacceptable consequences for the philosophy of mind. Given that no philosophy can coherently fail to give a satisfactory account of ourselves as feeling and thinking agents (what else are we doing when we do philoso-

phy?), we should take the disastrous consequences of mechanism for the philosophy of mind as very good reason for rejecting not *just* naturalism, but mechanism in general as a philosophy of nature.[1] Descartes was correct to accept dualism on the assumption of mechanism in nature, but if a viable alternative to mechanism were available, we could then avoid dualism and all of its problems.

In what follows in this chapter, I will outline the philosophy of nature, *Aristotelian hylomorphism,* which is historically mechanism's most significant competitor. I make no pretense to offering a compelling defense of this philosophy of nature. My intent is merely to propose it as an alternative that may well avoid the difficulties with which we are saddled by mechanism. A worthy defense of this view far outstrips what I can achieve in this book. Nevertheless, our foregoing discussions have shown us that progress in the philosophy of mind is unlikely until we make progress in the philosophy of nature. If you have begun your own philosophical inquiry with the philosophy of mind, then my recommendation to you is to leave the philosophy of mind aside for the moment, and begin an investigation into the philosophy of nature, wherein you may evaluate Aristotelian hylomorphism on its own grounds. In the next chapter we will discuss the fruits of this philosophy of nature for the philosophy of mind.[2]

1. In his recent book, *Consciousness and Mental Life* (New York: Columbia University Press 2008), Daniel N. Robinson outlines in great detail a post-mechanist strategy similar to what I am recommending here. See also Thomas Nagel, *Mind and Cosmos* (New York: Oxford University Press, 2012).

2. The term "philosophy of nature" is one foreign to contemporary Anglo-American philosophy, though it is quite common in the Thomistic tradition. Contemporary analytic philosophers will recognize what I call the philosophy of nature in this chapter as a species of metaphysics pertaining to causation, identity through change, and material constitution. For an introduction to how these views are dealt with by Thomistic philosophers see Leo Elders, *The Philosophy of Nature of St. Thomas Aquinas: Nature, Universe, Man* (New York: Peter Lang, 1997); Clarke, *The One and the Many: A Contemporary Thomistic Metaphysics;* Jacques Maritain, *The Philosophy of Nature* (New York: Philosophical Library, 1951), and *The Degrees of Knowledge* (Notre Dame, Ind.: University of Notre Dame Press, 1999), 184–214. E. J. Lowe provides an excellent introduction to the contemporary discussion of these issues in *A Survey of Metaphysics.* Though there is a sometimes

The Problem of Change

Among ancient Greek philosophers the problem of change is a primary issue, one that needs to be dealt with before making progress on any other philosophical concern. In this section we will discuss this problem and outline the principles necessary for its resolution. In the following section we will consider one the most common attempts, among both ancient and contemporary philosophers, to resolve it, *atomism,* though we will see that atomism is subject to very powerful objections. Assuming that atomism is unacceptable, we will then turn in the next section to another approach to the problem of change that was common among ancient and medieval philosophers and is once again being taken seriously by a minority of contemporary thinkers, Aristotelian hylomorphism.[3]

---

bewildering disparity in terminology and philosophical methodologies between these two streams of discussion, my own view is that the problems and solutions entertained by the Thomistic and analytic traditions of philosophizing are ultimately commensurable.

3. Obviously, given its name, this theory originates in the writings of Aristotle, with particular reference to the problem of change in his *Physics* and *On Generation and Corruption,* in *The Complete Works of Aristotle: The Revised Oxford Translation,* ed. Jonathan Barnes (Oxford: Oxford University Press, 1984). References to Aristotle will be given by the title of the work and line numbers in the Barnes volumes. St. Thomas Aquinas adopts Aristotle's views on these matters, and his most sustained treatments of these issues (with particular emphasis on the problem of change) are *On the Principles of Nature,* in *Aquinas on Matter, Form, and the Elements: A Translation and Interpretation of the De Principiis Naturae and the De Mixtione Elementorum of St. Thomas Aquinas,* ed. and trans. Joseph Bobik (Notre Dame, Ind.: University of Notre Dame Press, 1998), and *Commentary on Aristotle's Physics,* trans. Richard J. Blackwell, Richard J. Spath, and W. Edmund Thirlkel (Notre Dame, Ind.: Dumb Ox Books, 1999). My discussion will follow Aquinas's position fairly closely. Whether or not Aristotle and Aquinas are of the same mind in all of the details of these and related issues is a matter of some scholarly debate, though we can leave those interpretive issues aside for our purposes. The need for brevity requires me to ignore the primary competitor to Aristotle's solution to the problem of change, Plato's theory of the forms. See Jacques Maritain, *Introduction to Philosophy* (New York: Rowman and Littlefield, 2005), and E. Gilson, *Being and Some Philosophers* (Toronto: Pontifical Institute for Medieval Studies, 1952) for a good introductions to the contrast between Platonic and Aristotelian views. There are philosophers of mind who defend hylomorphism, though they do so from a decidedly non-Thomistic perspective and often independently of the historical figures who defend such views. These versions of hylomorphism are certainly worthy of consideration, but in what follows I will use hylomorphism in reference to more traditional views. See William Jaworski, *The Philosophy of Mind: A Comprehensive Introduction* (New York: Blackwell, 2011), especially chs. 10 and 11.

It is empirically obvious to us that natural objects change according to stable patterns, and indeed some of these patterns are seemingly inexorable, for example, "What goes up must come down!" In other words, natural change happens in a stable and predictable way. For example, when I release a pen it always falls to the ground, rather than turning into a winged monkey and flying away; or when I plant an acorn, I normally get an oak tree, and not a bicycle. Let's call these well-established patterns of change attested to by our universal common experience *laws of nature.* Moreover, these law-governed series of changes involve two different kinds of change. Consider a normal cat, Fluffy. Suppose for fun, the neighbor kid paints Fluffy blue. Fluffy has certainly changed; he was once calico and now he is blue. Notice, however, that this change in Fluffy is limited, because Fluffy has survived the change; Fluffy is still the same individual cat that he was before being painted blue. We call this an *accidental change,* because it involves a change of an *accidental property,* and by "accidental property" we mean a nonessential aspect of something or a way something exists that does not define it in terms of a fundamental kind of thing, for example, *being blue* as opposed to *being a cat.*[4] Instances of Fluffy's moving around, growing, shedding hair, losing one of his paws, and the like, are all cases of accidental change.

Suppose that Fluffy carelessly stands a bit too close to a wood chipper, and is thereby pulled through the machine. Fluffy has definitely changed, but this is no accidental change: whatever the essential properties of being a cat might be, the four pounds of

4. Do not confuse accidental properties with contingent properties. Accidents are nonessential or nondefinitive aspects of things, but there are cases of necessary (noncontingent) accidents. For example, it is a necessary accident of human beings that they are risible, that is, given to laughter, but this is not the essence of human beings. We would do better to define *humanity* in terms of *rational animality.* Traditional logicians typically call necessary accidents that are caused by the essence of a substance properties (*propria*). I, however, have followed the contemporary terminology in using "property" to refer to any characteristic of a thing, whether contingent/necessary, accidental/essential. For an excellent introduction to and defense of the traditional approach to the distinction between property and essence, see Oderberg, *Real Essentialism.*

blood, bone, hair, and other bits ejected from the wood chipper certainly lack them. Fluffy does not survive the change, because when the essence of being a cat is lost the individual cat is lost too; the remains at the far side of the wood chipper are no longer Fluffy, nor do they amount to a cat. We call this a *substantial change,* because it is not merely a change in the accidental characteristics of the substance, but *a change in the substance itself.*[5] Moreover, Fluffy's demise in the wood chipper is a specific type of substantial change we call *corruption,* whereas Fluffy's original coming to be with the union of a sperm and ovum is a case of *generation.* Suffice it to say, our experience is replete with apparent examples of generation, corruption, and accidental change. None of this seems philosophically earth-shattering, but we will see in a moment that there is a problem looming in the background.

The vast majority of philosophers accept the principle of noncontradiction, that is, something cannot both be and not be in the same respect at the same time. For example, something cannot both exist and not exist at the same time, nor can it both be blue all over and not blue all over at the same time. The principle of noncontradiction is a basic first principle, in the sense that it is a self-evident truth that cannot be inferred from any more basic principle. Indeed, it is the principle we need to get all reasoning off the ground, so there is no way it can be rationally defended by any other principle. To deny the principle of noncontradiction, however, is to fail to make sense.[6] Parmenides, one of the great grandfathers of ancient Greek philosophy, introduces the

5. Traditionally, essential characteristics are said to be in the logical category of substance that defines something as member of a fundamental kind. Thus, a substantial change is a change within that category, whereas accidental change is a change within the accidental (nonessential) categories.

6. Aristotle famously makes this point in *Metaphysics,* 1005b1–1011b23. Also note that one should not interpret the principle of noncontradiction as though it governed what we *believe* as opposed to what *is actually the case.* For example, the principle of noncontradiction does not preclude that "Martha believes that the ball is completely blue" and "Patrick believes that the ball is not completely blue" can both be true in the same context. Per the principle of noncontradiction, two people can have contradictory beliefs; we just know that one of them is wrong (even if we don't know which). On the other hand, the prin-

problem of change by following what he takes to be the consequences of the principle of noncontradiction for change. In fact, Parmenides argues that this principle implies that neither accidental change nor substantial change is possible. The following is the sort of reasoning that leads Parmenides to such a conclusion:

(1) If changes occurs, then the *non-P* becomes *P.* (premise)

(2) It is impossible that the *non-P* become *P.* (premise)

Therefore:

(3) It is not the case that change occurs. (from 1 and 2)[7]

Premise (1) is without a doubt broadly correct, though we will later see some need for qualification. Indeed, if there were no difference subsequent to the change, there would be no sense in which a change occurs. We might put this by saying that change must involve actual coming to be; whatever else change involves, it must include something new. Even in negative changes, for example, the *blue* becoming *non-blue,* there is still a new way of being introduced, a non-blue actuality. For now let's just call the entity that accounts for the coming to be in change the *principle of actuality,* while leaving aside what such a principle is in any particular case. One might retort that *non-P* doesn't become *P* in a change, but then it seems that we don't have change, but replacement, that is, one thing going out of existence only to be replaced by something distinct.[8] Parmenides maintains that premise (2) is

ciple of noncontradiction does deny that "The ball is completely blue" and "The ball is not completely blue" can both be true in the same context. Failure to attend to this distinction sometimes leads people to deny the principle of noncontradiction, though this confusion is easily avoided.

7. This is a simplified version of the Parmenidean argument, and there is much scholarly debate regarding exactly what Parmenides is up to in the relevant texts. For a detailed reconstruction of Parmenides's position, see Patricia Curd, *The Legacy of Parmenides: Eleatic Monism and Later Presocratic Thought* (Princeton, N.J.: Princeton University Press, 1998). For excerpts from Parmenides's original texts, along with other ancient irrealists about change, see *The Presocratics Reader,* ed. and trans. Patricia Curd and Richard McKirahan (Indianapolis, Ind.: Hackett Publishing, 1996), 29–56.

8. This is part of what motivates another famous ancient irrealist regarding change: Heraclitus, who believed that what looks like process of change by material substances is actually the replacement of one substance with another, and therefore not literally change. See Curd and McKirahan, *The Presocratics Reader,* 29–56.

a consequence of the principle of noncontradiction. His point is that if *non-P* became *P,* then *non-P* would be *P,* which is clearly a contradiction. Thus, given (1) and (2), Parmenides validly concludes that change, both accidental and substantial, is impossible.

The Parmenidean argument can at first be worrisome, because it seems to show that the unreality of change follows from the very foundation of all rational thought, the principle of noncontradiction. Notice, however, that even though it is not a logical truth, the fact that change occurs is likewise something that we should take as a first principle. If anything is evident given our experience, certainly change is. Indeed, even radical skeptics who doubt the existence of an external world or even a substantial self, seldom doubt the reality of change. Thus, we should not be ultimately troubled by Parmenides's argument, because given the status of the principle of noncontradiction and the evident fact of change, we know there is some error somewhere embedded in Parmenides's reasoning. That is not to say, however, that we should not take this argument seriously, because it certainly shows us the need for philosophical subtly in explaining what change really amounts to (even if we do not need such subtleties to know that changes is real).

To see the confusion motivating Parmenides's argument, consider the two ways of interpreting the consequent of (1) and the proposition said to be impossible in (2):

(4*) For some property, *P,* the lack of *P* becomes *P*.

(4**) For some entity, *x,* and property, *P, x* wasn't *P* (*x* lacked *P*), and *x* became *P* (*x* gained *P*).

Parmenides is correct when he claims that (4*) is impossible—the absence of a principle of actuality cannot become that actuality—but he is incorrect to think that (4*) represents what happens in change. That is, the things involved in change are not just a principle of actuality and its absence, but entities that lose absences and gain actualities. For example, when Fluffy turns blue, it's not the lack of blue that becomes blue, but *Fluffy,* who

was previously lacking blue, that becomes blue. Notice that (4**) represents this fact, as it claims that something undergoes the change—the change is not just a relation between an absence and an opposed actuality but the actuality and *something* that lacks it. That is, something that preexists the change comes to be in a new way. Thus, we should interpret (1) in terms of (4**):

(1*) If change occurs, then for some *entity*, *x,* and property, *P*, *x* wasn't *P* (*x* lacked *P*), and *x* became *P* (*x* gained *P*).

If we do not interpret (2) in terms of (4**), then the argument will be invalid, so we need to recast (2) as

(2*) It is impossible that *x* wasn't *P* and *x* became *P.*

The problem now is that (2*) is unsupported by the principle of noncontradiction and contrary to the evident fact of how change works in nature. Thus, we can conclude that Parmenides's argument rests on a failure to recognize that there is another principle involved in change in addition to the principle of actuality; that is, there is something that gains the form through the process of the change.

We call this substratum of the change the *principle of potentiality*, because, even though it lacks the actuality, it must obviously be such that it has the potential for the actuality. Thus, the prior lack of the principle of actuality in the principle of potentiality cannot be an essential absence, but is a way of being that the substratum can have under certain conditions. That is, we must make a logical distinct between a *privation* and a *negation,* and this is just how Aristotle frames the issue in his *Physics.*[9] Consider the two following propositions:

(5) The rock is blind.

(6) William is blind.

Let's take "blind" simply to mean *the lack of the power of vision.* Proposition (5) is a case of negation, as it denies that a certain in-

9. See Aristotle, *Physics,* 189b30–191a22. Aquinas also applies this distinction to the problem of change in *Principles of Nature,* §§6–8.

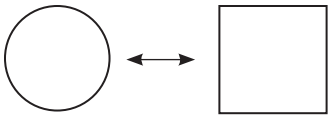

FIGURE 7-1. PRINCIPLES OF CHANGE

dividual has a property that is excluded by the essential characteristics of that kind of substance. Rocks are not the sorts of things that could have the power of vision in the first place; they lack the *natural potential* for having vision. Proposition (6), however, is a case of privation, because the power of vision is a property appropriate to the nature of a human being under standard conditions. William is the sort of being that has a natural potential for vision, and in fact, under normal conditions human beings develop vision as a matter of their standard pattern of change. Thus, blindness in William is not merely a *negation* of sight, but a *privation* of sight. The absence of a form prior to a change must be a privation (not a mere negation) in the substrate for such a change to be possible. Otherwise, change would proceed from a substrate with no potential for such a change, which would be impossible.

These points can be better understood by considering figure 7-1. Suppose that the double arrow represents the fact that a certain object can change back and forth between *being square* and *being circular.* If the circle just were *circularity* or *circularity* were *essential* to it, then it would be impossible for it to become a square. If, however, there is a substrate that has *circularity* as an accidental property while also possessing the potential for *squareness* (*though not simultaneously*), then a change from being a circle to being a square would be possible, and the same can be said for a change from *squareness* to *circularity.* The point then is that the possibility of change presupposes three principles, a substrate or *principle of potency* (that which preexists and survives the change), a *principle of actuality* (the way of being that the substrate gains), and a *priva-*

tion (a lack of a characteristic that is still within the range of potential for the substrate). Once we make these distinctions, resolving the problem of change then amounts to giving an account of what the principles of potentiality and actuality are.[10]

Atomism and the Problem of Change

So what is the principle of potency in change? One of the common answers in the ancient world was provided by the atomists, such as Leucippus, Democritus, and Epicurus.[11] On this view, an *atom* is the smallest possible physical substance, something that itself cannot be divided in any way into something simpler. The ancient atomists are materialists, who believe that atoms only have the properties of size, shape, motion, and location, and all that exists are atoms and aggregates of atoms.[12] Atoms themselves are both incorruptible and ungenerated; those that exist have always existed and always will exist, and that is all there is. Indeed, as the ultimate substrate of change, atoms are what account for the generation and corruption of the substances we are acquainted with empirically. Atomism, though very old, is implicit in the materialism common among many contemporary philosophers. Certainly, whether there is in fact a smallest physical particular is something that most contemporary philosophers take as an empirical matter to be resolved by scientific investigation, and the going physical theories tell us that physical particles are not in fact incorruptible in the way they were thought to be in the ancient world. Nevertheless, many contemporary philosophers are atomists in the sense that they believe that physical

10. Notice that privations are not entities, strictly speaking, but absences of actualities that could be had by a certain substrate. We might say that the privation is a necessary principle of change, but always as an aspect of or through the substrate. Thus, we don't give an account of them independently of the substrate of which they are an aspect. See Aquinas, *The Principles of Nature,* §2.

11. For Leucippus and Democritus, see *The Presocratics Reader,* 79–88. For Epicurus, see "The Letter to Herodotus," in *The Epicurus Reader.*

12. Epicurus, however, includes void or space that atoms occupy as part of his ontology. See *Letter to Herodotus.*

objects are structured aggregates of physical particles (divisible or otherwise) that persist as distinct entities when they enter into a composition. Whether in its ancient or contemporary guise, atomism implies that change can be analyzed in terms of accidental changes of atoms, literally changes in their position with respect to each other. For example, Fluffy's becoming blue is ultimately just a rearrangement of atoms, as are Fluffy's generation and corruption. Collections of atoms serve as the substrates of change that have certain arrangements as privations, and their rearrangements are the actualities gained in change. Thus, all change is merely atoms actualizing their potential for various arrangements. On this view, the principle of potential that is needed to assuage the problem of accidental change is provided by Fluffy's atoms. Thus, when Fluffy becomes blue or loses his tail, we merely have a rearrangement of preexisting atoms that have a natural potential of combination and separation. Consider figure 7-2.

For the atomist, Fluffy just is an aggregate of arranged atoms, and suppose that the F-arrangement of atoms amounts to Fluffy *being calico,* and further suppose that his G-arrangement amounts to Fluffy *being blue.* This accidental change seems nonproblematical because the atoms serve as the principle of potential and continuity, and any sort of structural arrangement seems to be something within the natural potential of a sufficiently large collection of atoms. That is, the atoms, though they were F-arranged, had the prior potential to be G-arranged. Since, according to the atomist, Fluffy just is a collection of variously arranged atoms, it seems that he can very easily be seen to undergo accidental change.

The atomist will attempt to address the problem of substantial change in a similar way. Consider figure 7-3.

Suppose that C-arrangement is a structural arrangement of atoms sufficient to be cat. In this case, Fluffy was nicely C-arranged at T^1 and thereby constituted as a cat, but upon the dissipation of his atoms by the wood chipper at T^2 he suffered a

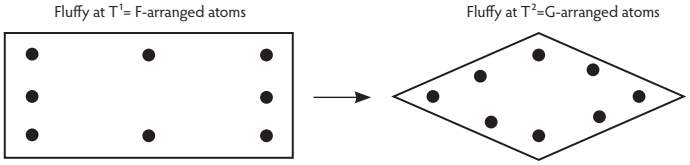

FIGURE 7-2. ACCIDENTAL CHANGE FOR ATOMISM

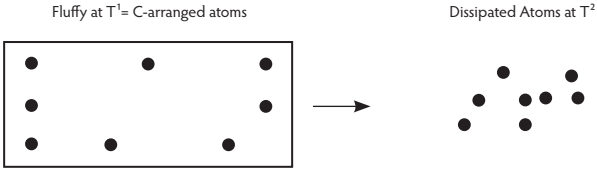

FIGURE 7-3. SUBSTANTIAL CHANGE FOR ATOMISM

sudden loss of *catness* because of the loss of the C-arrangement. If these very atoms could be arranged "C-ly" once again, then presumably Fluffy could be retrieved (though that is a controversial claim). The point here is that the essential nature or substance of a physical object is a certain arrangement of its parts, and generation and corruption are accounted for in terms of the loss or gain of such arrangements. Thus, all change is merely atoms actualizing their potential for various arrangements, a kind of accidental change in relative position for the atoms.

At this point, we should raise an objection to the atomist account of change. In particular, the atomistic account works only because the atomists were happy to accept something like a materialist account of substances like Fluffy (ultimately cats just are structured aggregates of atoms or are supervenient on such aggregates). Some of the arguments against materialism we discussed earlier in this book should lead us to be deeply skeptical about these claims. Fluffy clearly has properties, for example, sensations, which are not reducible to or strictly supervenient

on properties of his atoms. Since a cat is not straightforwardly identical to or supervenient on the atoms that compose it, when Fluffy comes to be, something substantially different from the atoms that compose him comes to be, but an atomist cannot account for this fact. At the end of the day the temptation for the atomist is to conclude finally, à la the eliminative materialist, that all that exists is just the atoms, but then it seems that we have once again lost the reality of the change through which Fluffy is generated and corrupted. As Yves Simon puts it, "For the materialist, then, change is more apparent than real. Underneath there is a big unchanging ball, the Parmenidean sphere, or there are little tiny balls, the atoms of Leucippus and Democritus," because either way, "you have a subject that does not really change and you arrange it in indefinitely many ways."[13] Thus, since atomism cannot account for the fact that Fuffy, *a cat,* came to be (and that seems once again to raise the specter of eliminativism), we should conclude that it fails to solve the problem of change.[14] Of course, an atomist might opt for an emergentist account of things like cats, wherein substantial changes occur when a fundamentally new kind of being emerges from a properly structured aggregate of atoms, but, as we have discussed earlier, it is far from clear that such an account is plausible given the broader mechanism we would expect an atomist to accept.

Even if we grant that something like a cat could be identical to, supervenient on, or emergent from a collection of atoms, problems similar to the problem of the many we discussed in chapter 2 still arise.[15] Let Fluffy be identical to, supervenient on,

13. Yves Simon, *The Great Dialogue of Nature and Space* (Notre Dame, Ind.: St. Augustine's Press, 1970), 30. This book is highly recommended as an introduction to and defense of the Aristotelian understanding of nature.

14. Not surprisingly, whether a materialist-atomist account of the continuity of a physical object can be given is certainly one of the most controversial issues in philosophy. For reasonably accessible, recent treatments of these issues see Merricks, *Objects and Persons;* E. J. Lowe, "Form without Matter," and Michael Rea "Sameness without Identity: An Aristotelian Solution to the Problem of Material Constitutions," in *Form and Matter,* ed. David Oderberg (Oxford: Blackwell, 1999), 1–22, 103–14.

15. In what follows I am adopting and developing a famous example originally intro-

or emergent from an aggregate of persisting atoms arranged *catwise,* S^1. Call the subset of S^1 that contains the aggregate of atoms arranged *catwise* that composes Fluffy minus his tail S^2. Notice that S^1 and S^2 are nonidentical (they are different aggregates of atoms), but both are sufficient, according to an atomist, to constitute a cat (whether you prefer identity, supervenience, or emergence).[16] Thus, we should conclude that there are in fact two distinct, but overlapping, cats occupying the space that outlines Fluffy. Let's call the cat identical to, supervenient on, or emergent from S^2 Luffy. There is, however, another aggregate of atoms, S^3, in the vicinity whose members are arranged *catwise,* but minus Fluffy's front left paw. Since S^3 is distinct from both S^1 and S^2 and sufficient for a cat, it now looks like we have three overlapping cats occupying the space that outlines Fluffy. Though three cats would be odd enough, things are even more crowded than that, because there is a distinct *catwise* arrangement of atoms for every atom that might be removed from S^1; none of these atoms

duced by Peter Geach in *Reference and Generality* (Ithaca, N.Y.: Cornell University Press, 1980). The problem we are about to discuss is one aspect of the problem of material constitution, about which there is a vast contemporary literature, and my remarks to follow would be taken as quite controversial. There are certainly a number of avenues to dodge these criticisms entertained by contemporary philosophers. My point is not to give a definitive refutation of atomism, but only to show that certain problems arise for it, which do not arise for hylomorphism. Of course, I think that these considerations are very good reasons for rejecting atomism, but I'm not trying to make any so strong of a case here. Also note that in chapter 2 we introduced the version of the problem of the many developed by Peter Unger, but Unger is not terribly sympathetic to the version introduced by Geach; see Unger, *All the Power in the World,* 366–71. The consideration of this and related problems has a long history, beginning in the ancient world and continuing today. For a good introduction to this area of contemporary metaphysics, see Lowe, *A Survey of Metaphysics,* 66–79; and Brian Weatherson, "The Problem of the Many," *in The Standford Encyclopedia of Philosophy,* http://plato.stanford.edu/entries/problem-of-many/. See also Michael Rea, *Material Constitution: A Reader* (New York: Rowman and Littlefield, 1996) for a collection of more technical papers. Roderick Chisholm's *Person and Object* is one of the most widely influential modern treatments of these issues. Trenton Merricks, in *Objects and Persons,* attempts to avoid these problems for atomism by eliminating composites like cats, whereas Peter van Inwagen attempts to preserve the reality of living things, while eliminating things like artifacts, in *Material Beings* (Ithaca, N.Y.: Cornell University Press: 1990).

16. Notice that we are assuming that the aggregates of atoms are arranged *catwise* in S^1 and S^2. Certainly, the aggregates alone, without the *catwise* arrangement, are insufficient for a cat.

is individually necessary for a cat, so any catwise arrangement missing just one of the billions of atoms in S^1 is sufficient to constitute a distinct cat. Thus, it seems that there are billions of overlapping cats occupying the space that outlines Fluffy. Certainly a legion of cats, in fact billions of them (and of course we could run this same argument for any physical object composed of atoms), where we think we see only one is enough for us to question whether atomism is in fact a viable option.[17]

Suppose that this proliferation of cats is not something you find particularly troubling. A related problem still lingers. Let's say, once again, that Fluffy stands a bit too close to the wood chipper, but this time he is not so unfortunate as to be sucked through, but he does lose his tail in the ordeal. Assuming that both Fluffy and Luffy existed prior to the loss of the tail, it seems that we have to choose from the following options:

(7) Fluffy was destroyed, but Luffy persists.
(8) Luffy was destroyed, but Fluffy persists.
(9) Both Fluffy and Luffy were destroyed, and a new distinct cat now exists in place of them both.
(10) Fluffy and Luffy became identical.

Proposition (7) is unacceptable, because the loss of a nonessential part, such as a tail, is certainly not sufficient for the destruction of an individual substance. Physical beings like cats are constantly gaining and losing nonessential parts, so accepting (7) would commit us to a great deal more generation and corruption than we can reasonably accept.[18] Proposition (8) is unacceptable

17. Not all philosophers are as dismissive of overlapping physical objects as I am in this paragraph. What is more, some philosophers even defend the notion that there are coinciding or co-located physical objects, i.e., distinct physical objects that do not merely overlap, but literally occupy the exact same space. Some defenders of these views might even be taken as members of the Aristotelian tradition (certainly broadly construed) that I am recommending in this chapter. E. J. Lowe offers an accessible defense of coinciding material objects in *A Survey of Metaphysics,* 68–72.

18. Some recent metaphysicians are willing to accept this consequence, because they accept a doctrine called *mereological essentialism.* That is, these philosophers argue that every proper part of a composite object (including its atoms) is necessary for its existence, so the loss or gain of any proper part amounts to a numerically distinct substance. G. W. Leib-

because S^2 still persists, and it is supposedly sufficient for Luffy, so we cannot conclude that we have lost Luffy. Proposition (9) seems to involve us in a needless generation of a third cat, and notice that we still have the atoms sufficient for Luffy hanging around. Proposition (10) has contradictory consequences, for example, the new Fluffy-Luffy identity would simultaneously have the property of having a tail at some prior time and lack the property of having a tail at that time. Of course atomists have produced subtle strategies for dealing with these worries, but these considerations serve to show that, even though it coheres well with many of our modern, mechanistic intuitions, there is some reason to look beyond atomism for our account of the possibility of both substantial and accidental change.

The Aristotelian Solution to the Problem of Change

On the face of it, the hylomorphist offers an account of change that is not altogether different from atomism. Both the atomist and the hylomorphist claim that change requires a substrate or principle of potentiality and a principle of actuality. For the atomist, the former is an aggregate of atoms and the latter is a structure or arrangement of those atoms. The Aristotelian calls these principles *matter* and *form.* Consider figure 7-4. Suppose the undifferentiated mass on the left-hand side actually becomes a circle, though it could become a square or a triangle. As we have seen, the principle of actuality is whatever it is about the result of the change that explains why it is actually a circle rather than a square or a triangle. The principle of potentiality is that which provides the potential for the change. The point then is that there is something present in the object that provides the potential to be in indefinitely many different ways, and we call this the *matter.* There is also something

niz's "The Monadology," in *Philosophical Papers and Letters,* trans. Leroy Loemker, 643–53 (Dordrecht, Netherlands: Kluwer Academic Publishers, 1969), is likely the most historically prominent defense of mereological essentialism, and see Roderick Chisholm's *Person and Object* for the most influential contemporary sympathetic treatment of the doctrine.

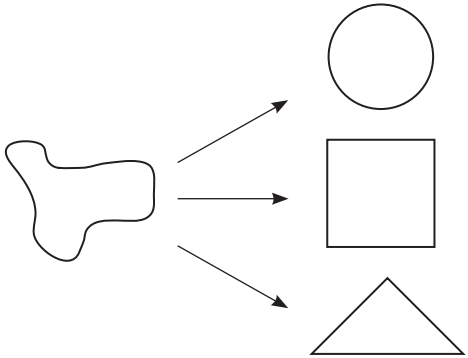

FIGURE 7-4. FORM AND MATTER IN CHANGE

present in the object that accounts for its being actually one way among its indefinitely many potentials, and we call this the *form.* In the diagram, the undifferentiated mass is the matter, and the various shapes it can become are the different forms of which it is deprived prior to the change.

Take the more concrete example of Fluffy's becoming blue. As we have seen earlier, there is certainly some difference between Fluffy before and after this change (he's now blue), so there must be some principle of actuality that accounts for this, that is, the way that Fluffy exists such that he is blue rather than some other color he might have been. Aristotelians call this principle of accidental actuality an *accidental form.* Fluffy's corruption by the wood chipper, however, is a substantial change. The pile of former cat parts is potentially a cat, but it is also potentially two small squirrels or several flowers (ultimately, Fluffy, the squirrels, and the flowers are all made from the same stuff). Whatever accounts for the difference between the remains of Fluffy and the former Fluffy is the *substantial form* of cat, just as the square shape accounts for the difference between the resultant figure and the shapeless mass. The point here is that any object that can

change, and this includes all material objects, cannot just be its matter (a material substrate), because that very same material substrate could become many other kinds of things. Thus, in addition to the matter of a substance, there must likewise be a substantial form that accounts for its coming to be the kind of thing it is, instead of the many different other kinds of things its matter might have been.[19]

So far, the atomist and the hylomorphist might appear to be in general agreement, inasmuch as they both recognize the need for matter, form, and privation (to use the preferred Aristotelian terms) as principles of change, but that appearance is superficial. This deeper disagreement is brought out well by considering the *prime matter* that many hylomorphists argue must be the principle of potential in a substantial change. Among Fluffy's various potentialities is *running* (*as opposed to eating, sleeping, playing, and so on*). Thus, when Fluffy begins to run, there is no problem, because the new accidental form was among his privations prior to the change. In this case, Fluffy, because he has a natural potential for running, is the matter of the change. In accidental change, the substance that undergoes the change is the matter.[20] Jeffrey Brower indicates why we cannot give the same account of substantial change:

> Unlike *sphericity* or *statuehood, humanity* is not the sort of form or property that can characterize its possessor accidentally. On the contrary, it is a form or property that characterizes its possessor essentially: if something is *human* at any time it exists, it must be *human* at all (possible) times it exists. For the same reason, Aquinas thinks, it makes no sense to speak of something *coming-to-be* human.[21]

19. See Aristotle, *Physics,* 190b15–22, and Aquinas, *Principles of Nature,* §6, for a clear statement of the need for three principles of change. See also George and Lee, *Body-Self Dualism in Contemporary Ethics and Politics, 66–67,* for a helpful statement of this point.

20. I am leaving aside for the moment the issue of whether the subject even of an accidental change can be an aggregate of persisting substances, which will be a point of disagreement between the atomists and certain hylomorphists.

21. Jeffrey Brower, "Matter, Form, and Individuation," in *The Oxford Handbook to Aquinas,* ed. Brian Davies and Eleonore Stump (New York: Oxford University Press,

The point here is that Fluffy does not have being a member of a different substantial kind among his natural potentials; for example, *Fluffy* cannot become three squirrels or a pile of moldering remains. In short, Fluffy has only cat potentialities, and non-cat substantial forms are therefore negations, not privations, in his case. The plausibility (though hopefully not the reality) of the wood chipper example demonstrates that Fluffy, however, is subject to substantial change in some sense; that is, we can begin with Fluffy and end up with something essentially different, but if Fluffy were the principle of potency for this change, we would face Parmenides's problem all over again. Thus, there must be a principle of potency in Fluffy that accounts for the possibility of substantial change, which hylomorphists call *prime matter.* In a substantial change, the result of a change is the loss or gain of a substantial form, for example, in Fluffy's case, whatever makes for the actuality of *catness.* Anything that has a substantial form has other substantial forms as negations, not privations, so prime matter must lack substantial form entirely; that is, it is the potentiality for any type of physical existence, but in itself, it is not an actual physical object of any kind. Furthermore, if Fluffy and the matter that ultimately composes him have distinct substantial forms, we would have to say that they are two distinct substances. We would then need to be substance dualists about anything that is subject to substantial change, for example, cats, trees, mushrooms. We can avoid these problems by recognizing that prime matter just is the pure potential to be in any one of indefinitely many physical ways with no intrinsic actuality of its own. As Aquinas puts it, "Since it has no form as an ingredient of its nature, prime matter does not have actual existence, since actual existence is only from a form. Prime matter exists only in potency."[22]

2012), 91. In recent correspondence Brower is now inclined to qualify the claim that all substances have their substantial forms necessarily (i.e., at all times during which they exist). This is certainly true of all material or natural substances, but theological considerations (e.g., the Incarnation) might lead us to believe that there are important exceptions.

22. Aquinas, *Principles of Nature,* §14.

Prime matter, something potentially any manner of physical existence though actually none, is admittedly a very strange notion, and even some of those who would count themselves broadly in the Aristotelian tradition have their doubts.[23] However strange prime matter might be, we are forced to accept something along these lines on pain of eliminating substantial change. That being said, with a bit of careful thinking, we can see that prime matter isn't *quite* so bad. The first thing to note is that prime matter and substantial form (*for the most part* in the case of the latter) do not exist independently of each other. They are what we call *co-principles* as opposed to *subsistent entities.* The latter are those individuals that can exist on their own—for example, the tree outside my window is a subsistent entity—whereas the former exist actually only inasmuch as they together form a compound substance. Prime matter and substantial form are co-principles of physical substances. One cannot trip over a pile of prime matter, nor can one pick up a substantial form independently from its matter, but one can pick up or trip over a compound substance composed of prime matter and substantial form. In ordinary physical objects, one cannot literally separate matter from form, any more than one can separate the spherical shape of a basketball from the rubber that composes the ball; there is some distinction between the shape and the ball, but the shape cannot exist without being the shape of some stuff, nor can the stuff ever exist entirely unshaped. Likewise, substantial forms of ordinary physical objects cannot exist without being the substantial form of some substance, nor can the prime matter of a substance *actually* exist without any form at all. Thus, though prime matter is prima facie strange, we have good reason to think there is such a principle; that is, it is the only way to account for substantial change in physical objects, and

23. For example, see John D. Kronen, Sandra Mennsen, and Thomas D. Sullivan, "The Problem of the Continuant: Aquinas and Suarez on Prime Matter and Substantial Generation," *Review of Metaphysics* 53, no. 4 (2000): 863–85, and Pasnau, *Thomas Aquinas on Human Nature,* 131–40. Brower attempts to answer these misgivings, at least as a coherent reading of Aquinas; see "Matter, Form, and Individuation."

the Aristotelian view does not entail that prime matter ever exists independently.[24]

One way to think of this issue is that if we were to count the individual objects in the universe, we would count neither quantities of prime matter, nor the substantial forms of ordinary physical substances; we would count Fluffy, but we would not count his prime matter and substantial form. Material objects are ultimately compounds of prime matter and substantial form, and it is material objects that we would count as individuals. Aquinas makes this clear when he argues that prime matter does not preexist God's creation of material objects.[25] It is not as if God takes a substantial form and literally puts it into a quantity of prime matter, as one might form a ball of clay into a cube. Rather, God creates physical substances, which are prime matter–substantial form compounds. Indeed, creating *actual* prime matter independently of its being in a compound substance is beyond the power of even God, because prime matter is not actually anything on its own. This is not to say, however, that prime matter and form are not real. The Aristotelian does not allow any ontological free riders in his philosophy of nature. If an entity plays a role in explaining an actual phenomenon, then we are committed to the reality of that entity. Thus, prime matter and substantial form are real, but they are real in the way that co-principles of compound substances are real, not in the way that individual substances are real.

We call the Aristotelian view that I have outlined thus far *hylomorphism,* literally from the Greek "matter" and "form," because on this view, physical substances are substantial form–prime matter compounds. It is important to emphasize for what follows that the compound substance, as constituted by inseparable (for the

24. John Haldane does much to demystify the notion of prime matter in an Aristotelian-Thomist philosophy of nature in *Reasonable Faith,* 18–36. See William A. Wallace, *The Modeling of Nature: Philosophy of Science and Philosophy of Nature in Synthesis* (Washington, D.C.: The Catholic University of America Press, 1996), 53–58, for discussion of the plausibility of prime matter (Wallace calls it *protomatter*) in light of recent particle physics. E. J. Lowe attempts to construct an Aristotelian view free of prime matter in "Form without Matter."

25. See *Summa Theologica,* Pt. I, Q. 44, A2.

most part) co-principles, is the basic entity of the Aristotelian ontology. In other words, the Aristotelian is anti-reductionist about physical substances, such that it is the substance and not its parts that is fundamental, and herein lies the most crucial difference with the atomist.[26] In fact, many Aristotelians defend the doctrine of the *virtual presence of the elements,* which basically claims that *a material substance does not have any individual substances as parts.*[27] Consider Fluffy eating some tuna, which presumably has quarks as its most basic parts, what we will call its *elements.* On this view, the quarks composing the tuna are prime matter–substantial form compounds, just like Fluffy in that respect. Aristotelians argue that when Fluffy digests the tuna, the elements of that substance no longer exist as substances in a strict sense; that is, their substantial forms are no longer present. If the quarks composing the tuna remained substances once incorporated into Fluffy, there would be multiple substances (indeed billions of substances as we saw earlier) occupying the space outlined by Fluffy's body. On the assumption that only one physical substance occupies the same space as Fluffy, and not indefinitely many things, we must then say that the substantial forms of the quarks are not actually present in Fluffy, and therefore the quarks are not present as individual substances. Rather, the Aristotelian argues, the *powers* of those substantial forms are present in Fluffy; the quarks are lost as substantial individuals, but Fluffy takes on their powers. Thus, the elements are present in the sense that their essential capacities have been adopted by the substance, what we call *virtual presence* as distinguished from *substantial presence,* but they strictly speaking do not exist as discrete parts of the substance. The elements "lose their independent status as substances and, at least for the

26. Note that I say the Aristotelian is anti-reductionist about physical substance, but not material objects generally. For example, according to the Aristotelian Fluffy is not reducible to his material parts, but artifacts such as tables, ships, coffee mugs, and natural objects such as stones or sand dunes are reducible to their proper parts.

27. The view I sketch here is defended by Aquinas in *On the Mixture of Elements,* in *Aquinas on Matter, Form, and the Elements,* 103–28.

time being, assume the status of 'virtual parts' of the new substance through the active and passive causal powers with which they endow the new substance."[28]

The ultimate composition of a material substance is its combination of substantial form and prime matter. There are many other levels of composition we can consider, for example, elements such as quarks, molecules, cells, organs, but these components are virtually, but not substantially, present in a unified substance.[29] According to the version of Aristotelianism I am suggesting, I do not have a heart as a discrete part of my body, but I have a heart-shaped region of my body, and similar claims can be made regarding my cells, molecules, and subatomic particles. (That is not to say that I do not have a heart, along with other "parts," only that my heart is not present in me as a discrete, individual substance.) The doctrine of the virtual presence of elements is probably the most difficult aspect of the Aristotelian position to swallow, especially given that the reducibility of physical substances to their "parts" is an unquestioned assumption for most contemporary philosophers. The doctrine is asserted, however, as part of a general philosophy of nature, of which it seems to be an inevitable consequence. As counterintuitive as it may be, the virtual presence of the elements is something one must accept, if one is going to accept the Aristotelian account of substantial change and the consequent account of physical substances as substantial form–prime matter compounds.[30]

28. A. J. Freddoso, "Suarez on Metaphysical Inquiry, Efficient Causality, and Divine Action," in Francisco Suarez, SJ, *On Creation, Conservation, & Concurrence: Metaphysical Disputations 20–22,* trans. and intro. A. J. Freddoso (South Bend, Ind.: St. Augustine's Press, 2002), xxxi.

29. William Wallace goes a long way toward showing how the various levels of composition within material substances revealed by modern science can be squared with the Aristotelian priority of the substance as a whole in *Modeling Nature,* 1–113. E. J. Lowe defends a basically Aristotelian position about physical substances while granting the substantial presence of physical particles by allowing for the co-location of physical substances; e.g., he would say Fluffy and the matter that composes Fluffy are both substantially present in the same place. See Lowe, "Form without Matter."

30. Patrick Toner does much to motivate the plausibility of a similar position in

Forms, Final Causes, and Scientific Explanation

Consider once again Fluffy's generation from sperm and ovum. Suppose we are not only concerned with what makes such a change possible, that is, the principles of change, but now we want to know why the change occurred, that is, the causes of change.[31] In other words, suppose we are not wondering, "What makes it possible for there to be change at all?" but, "What made Fluffy change in the way that he did?" Matter and form, though they are principles of change, are also important for answering questions regarding the causes of change. What something is made out of (matter) and what differentiates that thing from other types of substances made of the same kind of stuff (form) partially explain why such a thing comes to be. We call the former the *material cause* (because it pertains to *matter*) and the later the *formal cause* (because it pertains to *form*). Once Fluffy, or any other physical substance, has come to be, he follows a stable pattern of changes that are definitive of his kind; that is, Fluffy undergoes the changes characteristic of *catness,* but not *treeness* or *bumble-bee-ness.* Since Fluffy's prime matter, his ultimate material constituent, is equally disposed to *any* possible series of physical changes, our explanation as to why Fluffy undergoes only cat changes will be in terms of his substantial form. It is the formal cause, not the material cause, that explains why a natural substance follows a particular course of change governed by laws of nature. In other words, laws of nature are themselves only descriptions of the typical patterns of change natural substances undergo in virtue of their substantial forms.[32]

"Emergent Substance," *Philosophical Studies* 141, no. 3 (2008): 281–97. Another interesting defense of virtual presence can be found in Michael Stork, "Parts, Wholes, and Presence by Power: A Response to Gordon P. Barnes," *Review of Metaphysics* 62 (2008): 45–59.

31. See Aristotle, *Physics,* 194b16–195b30 and Aquinas, *Principles of Nature,* §§17–22.

32. This notion of a substantial form as a "law-maker" is elaborated and defended in Alexander Pruss, "What are Aristotelian Forms?" http://bearspace.baylor.edu/Alexander_Pruss/www/papers/Forms.html.

The material and formal causes, what Aristotelians call *the intrinsic causes,* don't make anything happen, so we additionally need *extrinsic causes* to have a complete explanation. If matter alone could make the change happen, there would be no need for an extrinsic cause, but in that case we would have no explanation as to why the matter realized one of its potentials at the expense of all others. The form does not exist in the substance until after the change, so it cannot be what makes the substance undergo the change.[33] We need, first of all, an *agent* or *efficient cause* that acts to bring about the effect. Somebody or something must *act* to move the matter from potentially having some form to actually having that form. In the case of Fluffy's coming to be, this will be his parents. In the case of Fluffy's corruption, the efficient cause is (at least in part) the wood chipper.

Suppose that anything, say a tortoise or a Goodyear tire, might just as well have resulted from the union of Mr. and Mrs. Fluffy Sr., or that Fluffy might just as well have shattered the teeth of the wood chipper under the same conditions. If that were possible, it would seem strange to say that Mr. and Mrs. Fluffy Sr. are the cause of Fluffy's conception or that the wood chipper caused Fluffy's demise, because there would be no intrinsic link or ordering between the efficient cause and the effect. The point to be gleaned here is that the wood chipper and Mr. and Mrs. Fluffy Sr. serve as agent causes only inasmuch as there is some ordering on their part toward a definite effect; that is, they are the kinds of things that under certain conditions bring about certain changes. These states toward which agent causes are disposed are what Aristotelians call *final causes,* and they explain why an agent cause brings about some particular effect.[34] In other words, the final cause is the end or terminus toward which the agent cause is directed.[35] Thus, a complete explanation of Fluffy's generation and

33. See Aristotle, *Physics,* 194b25, and Aquinas, *Principles of Nature,* §15.
34. See Aristotle, *Physics,* 194b32–195a2, and Aquinas, *Principles of Nature,* §16 and §22.
35. See Wallace, *Modeling Nature,* 15–18, for a helpful discussion of final causality, es-

corruption (and his accidental changes too) requires a fourfold account in terms of the material, formal, agent, and final causes of the change.

The foregoing presentation of Aristotelian hylomorphism has been, at best, cursory, and I have at critical junctures resorted to citing other authors and hand waving instead of making compelling arguments. Suffice it to say, I cannot defend a comprehensive view of nature here. My hope, however, is that you can see on the basis of this discussion that the Aristotelian-hylomorphic view is coherent and does solve important philosophical problems, even if I have fallen short of making a demonstrative case for it. Of particular note is that the problems of change and substantial unity that hylomorphism resolves are philosophical perplexities that arise independently of any worries regarding the philosophy of mind. The Aristotelian adopts hylomorphism not for the sake of the particular concerns of the philosophy of mind, but in order to explain the reality of change and physical substance. Notice that the Aristotelian eschews a mechanistic explanation of any material objects; substantial forms and final causes are irreducible realities necessary for understanding why things occur as they do in nature.

That being said, I do want to address, even if I need be cursory here too, one standard objection that is likely to occur to any reflective reader. The modern scientific method is often thought to succeed only at the expense of the Aristotelian view of nature. In modern natural science, we do not pay any direct heed to forms or final cause, but instead consider physical objects only in terms of their material and efficient (agent) causes. Physics and chemistry get along especially well at reducing things to their physical makeup (material cause) and considering the laws

pecially in terms of a terminus of change. For a detailed argument to the effect that final causality is an irreducible component of all explanation of natural change, see Stephen Mankin, "Aquinas, Natural Tendencies, and Natural Kinds," *New Scholasticism* 13, no.3 (1989): 253–74.

of efficient causality that govern the changes that such things undergo. What need is there for formal and final causes, as we can make sound predictions and build satisfying theories while availing ourselves of only matter and efficient causes? Aristotelian hylomorphism then seems to be a pre-scientific way of thinking that merely clutters our ontology with explanatorily sterile ideas such as *form* and *final cause.*

This objection is absolutely correct, if what the objector has in mind is the conduct of science. I don't think that the scientist should avail herself of anything more than material causes and laws of efficient causality, but that is not to say that substantial forms and final causes do not have much to contribute to our *philosophical understanding.* The problem, as Hillary Putnam puts it, is that "the notion of things causing other things is not a notion that is simply handed to us by physics."[36] In other words, it is a scientific question to ask after the types of physical objects there are, their physical composition, and the laws of efficient causality that govern their activities. It is a philosophical concern, however, as to what ultimately makes change possible and what ultimately grounds the laws of nature. The results of physics are compatible with any one of many different philosophical answers to these questions. On one hand, the physicist has as her goal the discovery of the broadest set of physical laws that predict and explain the changes of physical substances; she tries to discover what sorts of phenomena act as causes according to the laws of nature. On the other hand, the philosopher of nature is concerned with what makes change possible at all and why there are laws of nature at all. The former are questions for the physicist to address empirically, and when asking those questions there is no need to bring up forms and final causes. The latter are questions for the philosopher, and forms and final cause are relevant in this domain. Thus, there is

36. Putnam, *Renewing Philosophy,* 50. See also Popper and Eccles, *The Self and Its Brain,* 204–5. Ric Machuga makes a similar point while adopting an example from Putnam's text in *In Defense of the Soul,* 57–63.

nothing inconsistent in accepting both a classical philosophy of nature and modern physics.

Notions such as prime matter, final causes, and the virtual presence of the elements are strange to us, especially because none of these are things we can directly observe. As I argue in chapter 2 above, we make all sorts of inferences to unseen things when doing so is the best available explanation. Indeed, this is how we arrive scientifically at our beliefs in subatomic particles, dark matter, planets outside the solar system, and such; we infer the existence of such unseen (even essentially invisible) entities from their effects with which we are acquainted. The fact that the entities posited by Aristotelian hylomorphism are not directly observable should not scare us away.

Moreover, even though one can do science very well without forms and final causes explicitly in mind, it is notoriously difficult to account for the possibility of causal relations without them. David Hume, who is himself no friend of Aristotelian hylomorphism, argued that if all we have in our resources are the empirically detectable properties of physical substances and the prior patterns of change such substances have undergone, there is absolutely no reason to expect that such patterns will hold in the future.[37] That is, Hume showed that if we begin with the assumption of mechanism, we lose any necessity in the laws of nature (at least anything we would typically recognize as necessity). Why do these laws hold? The Humean would argue that however successful we are at discovering ever-broader laws of physics, ultimately there is no good reason as to why things follows such laws. On the Humean view, there is no reason (though our passions lead us to it nonrationally) for us to believe that tomorrow the law of gravity will not reverse itself. If one denies an explanatory role for anything nonempirical, as does Hume, then there is

37. Hume's clearest presentation of his skeptical arguments regarding causation is in *The Enquiry Concerning Human Understanding*, sec. 4.

no ultimate explanation of things. Nature might be one gigantic series of accidents that just happened to follow certain patterns in the past. In short, once we bar the notions of substantial form and final cause from our understanding of causality, we lose any account as to why things follow certain stable patterns of change and why any particular occurrence acts as a cause. Change just happens, and there is no reason for it (at least within our ken). That might be sufficient for doing science, but if we want an ultimate philosophical account of things, then we are going to need to go beyond such extreme empiricism. The point is that there is nothing pre-scientific or anti-scientific about Aristotelian hylomorphism. It is, at the end of the day, an attempt to explain philosophically what it is about physical substances (their substantial forms and final causes) that makes scientific explanation possible.

On the one hand, an Aristotelian is happy to allow a sort of methodological mechanism, wherein he grants that for scientific explanation, the physical properties, for example, mass, charge, velocity, of natural objects are sufficient for the explanation of their changes. On the other hand, if one desires an ultimate, philosophical explanation of change and material substance, one needs to go beyond the austerities of mechanism, because substantial forms and prime matter are required ultimately to explain the law-governed patterns of generation, corruption, and accidental change that physical substances follow. In short, the very possibility of mechanistic explanation requires that physical substances have a form-matter structure, which is itself not something that can be identified with these physical properties. Once again, I do not see this as a comprehensive defense of Aristotelian hylomorphism as a philosophy of nature.[38] I do hope, however,

38. For a detailed critique of attempts to given a non-Aristotelian account of causation, see Freddoso, "Suarez on Metaphysical Inquiry, Efficient Causality, and Divine Action," lxi–lxxii. A critique of Hume on causation can be found in Elizabeth Anscombe, "Causation and Determination," in *Causation,* ed. Ernest Sosa and Michael Tooley (Oxford: Oxford University Press, 1993), 88–104. See James D. Madden, "The Fifth Way, Scientism and Intelligent Design," 387–408, and "Leibniz and Malebranche on the Intelligi-

to have called into question the typical claim that Aristotelianism is merely a hangover from a pre-scientific understanding of nature. As we have seen in this chapter, the problem of change is a compelling philosophical concern that calls for an answer so that we can give a basic account of the reality of lawlike change in the natural world. Mechanism does not allow for a solution to this problem while accommodating the reality of many of the beings with which we are best acquainted (including ourselves). Aristotelian hylomorphism may use notions quite foreign to typically modern ways of thinking, especially the tendency to analyze everything in nature in terms of the categories of natural science. Hylomorphism is nevertheless worth taking seriously because it provides a solution to the problem of change, which is something we must do in order to account for the success of modern science. Otherwise we are left with a Humean skepticism even about the scientific method. In the following chapter, I will consider the implications of Aristotelian hylomporphism for the philosophy of mind.

---

bility of Nature," *Proceedings of the American Catholic Philosophical Association* 77 (2004): 173–188, for arguments to the effect that final causality cannot be eliminated from a coherent account of even efficient causation. For a highly technical defense of final causality in nature, see Robert C. Koons, *Realism Regained: An Exact Theory of Causation, Teleology, and the Mind* (New York: Oxford University Press, 2000); Michael Rea, though he does not defend an Aristotelian hylomorphic account in this text, argues against the coherence of any account of nature entirely devoid of the notion of purpose or final cause in *A World without Design.*

CHAPTER 8

ARISTOTELIAN-HYLOMORPHIC PHILOSOPHY OF MIND

In the previous chapter, I presented Aristotelian hylomorphism as an alternative to mechanism, which is an unspoken assumption among many contemporary philosophers. You will remember that in chapter 2 I presented mechanism as the view that the full account of physical objects is given by the physical properties of the atoms that compose them and the relevant physical laws. Thus, according to the mechanist, the activities of any material substance (including human bodies) can be fully explained in terms of the physical properties of the atoms (fundamental physical particles) that compose them. Our discussion in the previous chapter shows the deep disagreement between Aristotelian hylomorphism and mechanism. The Aristotelian argues that the problem of change requires us to posit substantial forms and final causes to explain the law-governed patterns of change that natural substances undergo, even if mechanical principles alone are sufficient for scientific purposes. That is, the Aristotelian argues

that we cannot give a completely mechanistic account of any physical substances that are subject to generation and corruption and follow natural patterns of change in general. Any such substance is a prime matter–substantial form compound; and it is the substantial form that is primary in determining the natural course of development of the substance. Thus, any attempt to reduce physical substances to the particles that compose them will inevitably leave out the most fundamental level of explanation.

As we turn to the implications of hylomorphism for the philosophy of mind, remember that the Aristotelian does not arrive at this view as an ad hoc attempt to gerrymander an account of nature around our commitments in the philosophy of mind. Rather, the Aristotelian takes this view in order to solve broader philosophical problems. In this chapter, I will apply hylomorphism to problems more directly relevant to the philosophy of mind and answer some of the most important criticisms of this position by recent philosophers.

The Souls of Living Substances

As you already know, the Aristotelian hylomorphist argues that all physical substances are ultimately substantial form–prime matter compounds. Aristotle himself was impressed with the fact that natural substances undergo standard series of changes ordered to certain ends that seem to define their kind. Fluffy the cat, ever since he was generated from sperm and ovum, sustains a consistent course of development through characteristic cat changes; for example, he grows, sheds his hair, chases mice, he is party to the generation of new cats. Aristotle calls the source of these self-directed changes the *nature* of the substance, and concludes that the nature of the substance must be its substantial form, because Fluffy has matter in common with things that do not undergo characteristically cat changes.[1]

1. See Aristotle, *Physics,* 192b10–193b20.

Natures can be arranged into a hierarchy. Elements, for the ancients earth, air, fire, and water, but for us fundamental physical particles, have a certain set of basic causal powers that dispose them toward combination into more complex physical entities. The substantial form of an element is that which must be present in addition to prime matter to actualize these causal powers. As higher levels of organization of elements occur, greater causal powers and fundamental capacities emerge. Eventually, self-sustaining compound substances with the power to take nutrition, grow, and reproduce—that is, living things—will emerge. The nature or substantial form of a living thing is whatever must be present to make something an *actual* organism of a certain kind instead of an aggregate of *potentially* living elements.[2] Aristotle calls the substantial form of any living being its *soul:* the soul "is an actuality of the first kind of a natural body having life potentially in it."[3] James Reichmann summarizes Aristotle's reasons for postulating a soul of all living things:

> Is there any way of explaining the living thing in terms of its parts? If there were, would it not follow that all bodies would be living? And would there then be any explanation whatever for the universal phenomenon of death? I submit that there would not. It was precisely to incorporate the common experience of living things as operative unities and the obvious phenomenon of limited life spans of all organisms that Aristotle postulated the existence of soul as the internal life source of living things.[4]

There is clearly a difference between the oak tree when it is alive and the nonliving remains used to make a piece of furniture. The former, and not the latter, grows, takes in nutrition, maintains and repairs itself, reproduces itself, and so on. There is some substantial form that accounts for the difference between the living tree and the potential tree matter. In light of this fact, Aristotle "viewed the soul as the first principle of life within all living

2. See Aristotle, *On the Soul,* 412a10–15 and *On the Parts of Animals,* 640b1.

3. Aristotle, *On the Soul,* 412a26.

4. James B. Reichmann, SJ, *Evolution, Animal "Rights," and the Environment* (Washington, D.C.: The Catholic University of America Press, 2000), 64.

things; that is, as the exclusive internal source of an organism's dynamic unity, of its ability to move and replicate itself and to perdure."[5]

We need to be careful with the talk of souls in all living things. On the one hand, Aristotle is perfectly serious about this. There is a difference between actually living and potentially living things, so there must be some formal principle that accounts for this difference. As I mentioned earlier, Aristotelians do not allow ontological stowaways in their account of nature. Souls, the substantial forms of living things, are therefore *real.* On the other hand, we cannot allow ourselves to be unduly influenced by substance dualist presumptions about the soul. Souls, in Aristotle's thinking, are not necessarily immortal entities capable of existing on their own. We should not be shocked by Aristotle's claim that every living thing has a soul, because by "soul" he means something quite pedestrian, namely the substantial form in virtue of which something is actually, as opposed to potentially, alive as a certain kind of organism. To take Aristotle's example, if an eye were a living substance in itself, then whatever it is that provides the actual *power of sight* would be the soul of the eye, or the "soul" of an ax (supposing that an ax were a living substance) is whatever it is that provides the actual *chopping power.* For the Aristotelian hylomorphist, the soul-body relationship is a special case of the form-matter relationship that pervades nature. That is, the soul and body are mutually dependent co-principles of an organism of a certain kind; they are not substances themselves, but constituents of a substance.

For the most part, soul and body are inseparable in the sense that one cannot exist without the other. Disperse the matter that composes the eye, and there is no principle of the *power of sight* that survives, and without the *power of sight,* the matter would not *actually* be an eye. Likewise, if one destroys the ax, there is no

5. Ibid.

principle of actuality of *chopping power* that remains, but wood and iron without the principle of *chopping power* fail to be an ax. There is nothing "spooky" going on here; the soul just is whatever accounts for something's being an actual organism. In some cases, the soul might just be a principle of organization such that when elements reach a certain level of highly complex arrangement a living substance results. There is nothing inconsistent between an Aristotelian understanding of the soul as the principle of actuality of a living thing and the insights of biology, as long as we do not presume that the latter entails materialism about living things, and there are plenty of reasons for doubting materialism with respect to conscious living things, as we have discussed above. The Aristotelian may then welcome whatever biochemistry has to tell us about the origins of life in terms of the material foundations of living things. There is no reason to doubt that the life sciences will one day be able to identify the physical processes through which living substances originally came to be, but that does not in the least show that to be alive just is to be a certain set of material constituents.

Aristotle notes, however, that some "parts" (powers or capacities) of a soul "may be separable because they are not the actualities of any body at all."[6] In other words, if there is some kind of organism whose characteristic powers simply could not be embodied in a physical object, that is, they could not be powers actualized in matter, then the soul of such an organism in some way acts independently of matter. In such a case, we would have good reason to conclude that the soul could exist without its matter. The point here is encapsulated in the famous scholastic dictum that *act follows on being;* if *x* acts independently of *y,* that is, *x* performs an activity that simply cannot be an activity of *y,* it follows that *x* exists independently of *y.* If there is some capacity that simply cannot be explained in any way as being embodied,

6. Aristotle, *On the Soul,* 413a5.

it follows that the soul of an organism that has such a capacity would in fact be separable from that matter. Such a soul could exist without any matter at all. At this point in our inquiry, we have no reason to believe that there is such a thing, because the capacities of nutrition, growth, and reproduction, such as we find them in plants (and other organisms), are not at all independent of material embodiment. The soul of a tree does not survive the lumberjack's handiwork.

The Souls of Conscious Substances

Return to our example of Fluffy the cat and his regrettable encounter with the wood chipper. There is certainly a difference between the living, breathing Fluffy and his post–wood chipper remains. The latter does not eat, grow, or reproduce, whereas the former does. Nutrition, growth, and reproduction are not the only interesting powers possessed by Fluffy that distinguish him from a cat-shaped corpse. Fluffy is also conscious; that is, he is the sort of thing that can have sensations. To be a living cat, as opposed to a potential cat, entails that an organism have, inclusively, the powers of living things in general (growth, reproduction, nutrition, etc.) and sensation, even if the organism is not immediately exercising these powers at the moment; for example, Fluffy might be asleep, injured, or at a preconscious state of physical development. Of course there are also characteristics of cats that distinguish them from other organisms capable of sensation, but for our purposes we will focus only on what distinguishes a conscious organism from a merely living thing.[7] Whatever it is in

7. There is debate among Aristotelians as to whether the difference between, say, a house cat and a panther is essential, requiring a distinct substantial form, or accidental, requiring only a set of accidental forms. The more traditional view, and most certainly the view of Aristotle and St. Thomas, is that house cats and panthers have different substantial forms. Other Aristotelians, sometimes impressed by the results of modern evolutionary biology, argue that there really are only four types of substantial forms among natural things, i.e., the substantial forms of the most fundamental physical particles, the substantial form of living beings in general, the substantial form of conscious beings, and the substantial form of human beings. Within each of these categories, there are only accidental

virtue of which something is *actually a living, conscious organism* is the *animal soul.*

As in the case of the soul of living beings, we should resist the temptation to make either too little or too much of the souls of conscious animals. The souls of conscious animals are real, and not merely ways of thinking about or predicting the behavior of certain organisms. The soul of a conscious organism is that which explains the difference between it and something lacking the capacity for sensation. Since the capacity for sensation is real, the souls of conscious organisms are real; hylomorphists do not allow the power of sensation to be an ontological free-rider. Animal souls, however, are inseparable from their matter. Sensations are processes completely dependent upon the physical processes of certain kinds of nervous systems, even if there is no strict relation of identity between the two. Robert George and Patrick Lee, who defend a broadly Aristotelian philosophy of mind, make this point clear:

> In the dog's case the sense organs naturally function as instruments of sensation, and the animal's consciousness seems to be in every way dependent on the bodily parts of the animal organism. Thus, it is more natural to suppose that sensation is a complex action, with changes in the sense organs, nervous system, and brain, being parts of that conscious, sensitive action, rather than being merely extrinsic (or preparatory) to it.[8]

Having a sensation, whether in a human or a nonhuman animal, is not identical to any neurophysiological process, but it is not external to such processes either. Sensations, we might say along with an emergentist like Searle, are systems features of neurophysiological processes, even though they are not strongly reduc-

---

differences between the "species." For a defense of this view, see Mortimer J. Adler, *The Difference of Man and the Difference It Makes* (New York: Fordham University Press, 1993) and Machuga, *In Defense of the Soul.* For simplicity's sake I will speak of only the four broad categories of substantial forms, though for our purposes we need not take a side in this debate, for now at least.

8. George and Lee, *Body-Soul Dualism in Contemporary Ethics and Politics,* 11. George and Lee make a strong case for the bodily nature of sensation, in both humans and nonhuman animals, throughout the first chapter of this book.

ible to such processes. Indeed, St. Thomas seems to agree with the emergentists that sensations emerge from neurophysiological organs and processes; Aquinas attributes a "spiritual" existence to such states, though he likewise argues that they are entirely dependent on the sense organs.[9] The soul that accounts for such a power is not separable from the matter of a conscious being. Animals that have developed certain types of nervous systems have the power to have sensations, and whatever distinguishes them as living beings with this power is the animal soul.

The fact that the Aristotelian hylomorphist takes sensations as, in some sense, emergent features may strike you as curious, as I have claimed that hylomorphism can provide us with a plausible philosophy of mind, but in chapter 6 I did much to call emergence theories of psychological states into question.[10] You will remember that I offered three lines of criticism against emergence views: (a) naturalist emergence theories seem to resort to a question begging appeal to brute facts; (b) the universal content of thoughts cannot be accounted for in terms of emergence of physical particulars; and (c) appeals to emergence do nothing to address the paradoxical consequences that plague all forms of naturalism with respect to intellectual agency. Since our concern at the moment is the supposed emergence of sensations and not thoughts and agency, neither (b) nor (c) need concern us now. We will, however, take the time to address (a) immediately.

In chapter 6, drawing on Strawson among others, I argued that the emergence theories often give the appearance of appeals

9. *Summa theologica,* Pt. I, Q. 78. A. 3.

10. If by "emergent," we mean that sensations are nonphysical though caused by physical processes without any direct supernatural or nonphysical agency, I think that all hylomorphists are likely to agree. If we mean that the power of sensation has an emergent origin from nonconscious living things, many hylomorphists will disagree, as they argue that no cause can be greater than its effect in perfection, so an unconscious being could not be the cause of a conscious being. See Aquinas, *On Truth,* Q. II, A. XIV; and David Oderberg, *Real Essentialism* (New York: Routledge, 2007). We do not need to take a stand on this issue here, but do note that I will argue in the following section that thought, and therefore the human soul, cannot be emergent in either sense. Thanks to Jean Rioux for raising this issue to me.

to brute fact. That is, given their mechanistic understanding of the physical constituents of neurophysiological processes, there seems to be no possible, or at least plausible, link between conscious states and their physical foundations. Without any sort of explanation, it is unclear why we would say that conscious states emerge from neurophysiological states. In short, the emergentist admits that conscious states are not the same kind of states as neurophysiological states, but she nevertheless argues that the former emerge from the latter. Without a plausible metaphysical model, this claim rings hollow. What then licenses the Aristotelian to appeal to emergence in the case of life and consciousness?

The fact that Aristotelian hylomorphism relies on a broader, nonmechanistic philosophy of nature makes all the difference. As one Thomist puts it, the Aristotelian conceives of "inorganic matter in such a fashion that, thanks to its substantial form, it could compound itself in the structure of organized beings. There is perhaps nothing purely material in nature."[11] We would expect a non-hylomorphist emergentist to believe that an organism just is a structured aggregate of *substantially present* microphysical particles, quarks, electrons, and the like, and he supposes that the capacity to have sensations emerges from such aggregates. Since the organism just is an aggregate of properly structured microphysical particles, the prior potential for the power of sensation must be provided for by nothing more than the actual quarks, elections, and so on. There is nothing about the essential properties of microphysical particles that in the least recommends the possibility of sensation, and this is why mechanistic emergence theories strain credulity. Since, as we discussed in the previous chapter, the Aristotelian hylomorphist believes that the living organism is composed of *virtually present* (not substantially present) microphysical particles, much of this difficulty can be avoided. The composing microphysical particles are present as powers

11. Etienne Gilson, *From Aristotle to Darwin and Back Again,* trans. John Lyon (Notre Dame, Ind.: University of Notre Dame Press, 1984), 115.

of the organism, but they are not present as substances or entities in their own right. The organism is fundamentally a compound of prime matter and substantial form, the former being the potentiality for physical being of all kinds, whereas the latter is the principle of actuality or limitation that determines the organism to be a substance of one kind in particular. Thus, it is not the potentiality of just the microphysical particles from which the power of sensation is derived, but the general potential of prime matter to be actualized by a certain type of substantial form. The point is that, for the Aristotelian hylomorphist, prime matter has the potential for the power of sensation (as it has the potential for all physical capacities), which is actualized with the substantial form of a conscious animal; "substantial forms preexist in the potency of matter."[12] Certainly, only organisms composed of microphysical particles have sensations, but remember that these particles do not strictly survive the composition, so it is not from their essential characteristics alone that the power is derived. Rather, the power is derived from the prime matter–substantial form composition, even though in the case of animals the result is necessarily an organism that has microphysical entities as virtual components. It is entirely consistent with the Aristotelian understanding of nature to argue that matter has a natural potential to become alive and ultimately conscious, because the Aristotelian does not define matter in contradistinction to such properties up front. Prime matter is the potential to be in all possible physical ways, and the realization of those ways of being is accounted for by the substantial forms matter might take. Among these ways that matter can be actualized is to be an organism with the power of sensation.[13]

12. Brian Leftow, "Souls Dipped in Dust," in *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons,* ed. Kevin Corcoran (Ithaca, N.Y.: Cornell University Press, 2001), 121.

13. Joseph Bobik notes a certain limited congruity between Searle's emergence view and Aquinas'shylomorphism in *Aquinas on Matter, Form, and the Elements,* 294–97. See also Bobik's discussion of the *eduction* of novel substantial forms from lower levels of formal organization on 228–41 of the same volume.

You will likely remember that I, also in chapter 6, chided the panpsychist's attempt to solve the emergence problem by positing consciousness as a fundamental feature of all physical particles, just like charge, mass, and the like. The worry is that this seems to be an obvious bit of gerrymandering of our philosophy of nature to fit a particular, in this case naturalist, story we want to tell in the philosophy of mind. The problem is that panpsychism is intrinsically implausible, and we would never accept it unless we really wanted to defend naturalism in the philosophy of mind. It is suspicious simply to redefine what we mean by "physical" so as to include consciousness in a "naturalist" theory of mind. Note, however, that the hylomorphist does not build her philosophy of nature around concerns in the philosophy of mind. The matter-form distinction and subsequent rejection of mechanism in general are developed to solve a variety of philosophical issues, that is, the problem of change, material constitution, persistence, and so on. Even if this distinction did no work for us in the philosophy of mind, there would still be very good reasons to accept hylomorphism. Thus, hylomorphism, unlike panpsychism, is not an ad hoc philosophy of mind. Furthermore, the hylomorphist does not claim that fundamental physical particles are conscious, but only that prime matter has the potency for the power of sensations that can be actualized under certain conditions. Of course prime matter is part of the composition of fundamental particles and conscious organisms alike. That is not to say that the former has a proto-version of the power of sensation, but only that it is composed of something that has the potential for such a power when it is coupled with the right sort of substantial form. The hylomorphist does not redefine matter to make room for sensation in a way convenient for his desired outcome, but brings an understanding of physical reality in general to bear on the problems of the philosophy of mind.

Sensation is emergent on, though neither reducible to nor strictly supervenient on, neurophysiology, because sensations

are not *strictly* physical states or events. In fact, St. Thomas argues that sensations, even in animals, are in some sense *spiritual:*

> Now, immutation is of two kinds, one *natural*, the other *spiritual*. Natural immutation takes place by the form of the immuter being received according to its *natural existence*, into the thing immuted, as heat is received into the thing heated. Whereas *spiritual* immutation takes place by the form of the immuter being received, according to a *spiritual* mode of *existence*, into the thing immuted, as the form of color is received into the pupil which does not thereby become colored. Now, for the operation of the senses, a *spiritual* immutation is required, whereby an *intention* of the sensible form is effected in the sensile organ. Otherwise, if a *natural* immutation alone sufficed for the sense's action, all *natural* bodies would feel when they undergo alteration.[14]

This passage is highly technical in its scholastic terminology, but it serves well as a summary of how sensation works according to the Aristotelian hylomorphist. Aquinas recognizes two fundamental types of change or coming to have a form ("immutation" as he is translated above), that is, natural and spiritual. When natural change occurs, a form is received into some matter in the way we encounter in ordinary physical occurrences, for example, a stone's gaining the accidental form of heat when it becomes hot or a canvas gaining the accidental form of red as it is painted. In these cases the substances enduring the change come to exist as hot or red in a physical manner. St. Thomas, not entirely unlike the causal theories of intentionality advanced by some functionalists we discussed in chapter 5, argues that sensation involves a change in the conscious organism caused by the object of its sensation, that is; a causal relation with a hot object causes the organism to have a sensation of heat, and a causal relation with a red object causes an organism to have a sensation of red. Unlike causal theories of intentionality, St. Thomas does believe that sensation must result in an intrinsic, spiritual change in the conscious animal. The sense organs (or as we would more likely

14. *Summa theologica,* Pt. I, Q. 78, A. 3.

think of it, the appropriate neurophysiological structures) *become hot* or *become red* when the organism has such sensations. The point here is that having a sensation is to be caused by the appropriate object not merely to be in just any state, but to be in a state of awareness of a quality of such an object.[15] On St. Thomas's rendering, having such an awareness is secured only by a sort of identity, what is often called a *formal identity,* that is, the sense organs and the object of sensation must have the *same* form . As St. Thomas points out, it would be absurd to say that the form of red or heat exists in exactly the same way in the sense organ as it does in the stone or on the canvas; otherwise we would have to claim that stones and canvases have red or hot sensations when they become red or hot, and we would have to claim that there are tangibly red objects and hot objects in the sense organ when an organism has sensations of red or heat. The form in the object of sensation has a natural or physical mode of existence, whereas the form in the sense organ has spiritual or nonphysical existence, and the latter is what accounts for the qualitative and intentional nature of sensation. As Aquinas puts it, something can have the power of sensation inasmuch as it can "receive something immaterially.... For the [form] of a stone is in the soul but not according to the act of existing which it has in the stone,"[16] or "the sensible form is conditioned differently in the thing which is external to the *soul,* and in the senses which receive the forms of sensible things without receiving matter, such as the color of gold without receiving gold."[17]

We should be careful not to make too much of Aquinas's claim that the forms of the objects of sensation exist spiritually in the organisms that have sensations. Anthony Kenny does much to explain this point as follows:

15. *On Truth,* Q.2, A. 3 and Q. 10, A. 4 contains some of Aquinas's most helpful discussions on this point.

16. Ibid., Q. 21, A. 3.

17. *Summa theologica,* Pt. I, Q. 84, A. 1.

> The word "spiritual" here is surprising. It is meant to make a contrast with natural, i.e., physical, changes; but Aquinas does not mean that anything ghostly or immaterial is happening. On the contrary he frequently emphasizes that the powers of the senses … do not transcend the world of matter and can only operate under the appropriate physical conditions. In this context followers of Aquinas have preferred the word "intentional" to the word "spiritual."[18]

Sensations are nonphysical or spiritual occurrences in the sense that looking at or touching the sense organs would not reveal the sensation of red or heat; the sense organ has the form of red or heat, but not in a physical way. Nevertheless, sensations do occur in physical organisms, and they are entirely dependent upon (embodied in) the neurophysiological process of such organisms, and this is why I have argued above that it is best to think of the Aristotelian hylomorphist as giving an emergentist account of sensation. Moreover, one should note that Aquinas does not propose this account of sensation as an explanation of *how* sensation works physiologically, but as a philosophical account of what sensation *is.* As Kenny puts it: Aquinas's account of sensation "is not intended to be an explanation of the nature of sensation: it is meant to be a philosophical truism to help us to understand, not to explain, what happens when an animal sees or hears."[19] Thus, Aquinas is open to whatever the empirical sciences discover regarding *how sensation works,* though in any event, on his account, what sensation *is,* is the formal identity of the sense organ with the object of sensation.

Finally, St. Thomas's account of sensation offers us some broader philosophical advantages. I recalled a bit earlier in our discussion of causal theories of intentionality in chapter 5. Such theories of intentionality have the virtue of accounting for the fact that sensation is a capacity that human beings share with other conscious animals, and they are correct to claim that intention-

18. Anthony Kenny, *Aquinas on Mind* (New York: Routledge, 1993), 34.
19. Ibid.

ality is a product of causal interaction between an organism and objects in its environment. The problem with the causal account is that mere efficient causation, understood mechanistically, does not secure intentionality; the fact that a sensation is caused by a certain property in an object does not entail that the sensation is *about* or *of* the object. Articulating a notion of *aboutness* sufficient to account for intentionality is one of the most difficult projects in the philosophy of mind, and as we saw in chapter 5, materialist theories of mind are not up to this task. Moreover, nonhylomorphist emergence accounts are inadequate because their mechanistic underpinnings are insufficient to do more than beg the most important questions. St. Thomas's account of sensation offers an understanding of *aboutness*—that is, sensations are about the objects of sensation because they are *formally identical* to the objects of sensation—but in doing so Aquinas treats sensation as a physiologically grounded capacity. This position then squares quite well with both the fact that sensations have an intentional aspect that cannot be strictly reduced to neurophysiology and the fact that our capacity to have sensations is inseparable from neurophysiology and environmental interactions.[20] The prospect of giving such an account of the intentionality of sensation does much to recommend the Aristotelian-hylomorphic approach, especially because the failure to provide a satisfactory solution to this problem is one of the most distressing shortcomings of contemporary philosophy of mind.[21]

20. Anthony Kenny gives a recent, critical discussion of St. Thomas on the intentionality of sensation in "Intentionality: Aquinas and Wittgenstein," in *Thomas Aquinas: Contemporary Philosophical Perspectives,* ed. Brian Davies (New York: Oxford University Press, 2002), 243–56. See also Roger Pouivet, *After Wittgenstein, St. Thomas* (South Bend, Ind.: St. Augustine Press, 2006), 43–71.

21. Many contemporary proponents of St. Thomas's philosophy of mind motivate the Thomistic position on the intentionality of sensations on the promise of the fruits it bears for constructing an account of sense knowledge that avoids the difficulties of skepticism about material objects. On this view one is not directly aware of an idea distinct from the object of sensation, but the object of sensation itself, which is something made sense of by the notion of the formal identity between the sense organ and the object. These epistemological concerns take us far afield of our concerns here, but there is a great

The Souls of Human Beings

The Aristotelian-hylomorphic understanding of the souls of living beings and the souls of beings capable of sensations can be said to be naturalistic in the sense that life and sensation are biologically dependent capacities had by natural beings that do not seemingly require a direct supernatural explanation, though, of course, the Aristotelian-hylomorphic account of living and conscious beings depends on a much more robust philosophy of nature than the mechanism that undergirds contemporary naturalism.[22] Consider again Fluffy the cat in contrast to William, a human being. William shares with Fluffy the requisite powers of life and sensation, but he also has the power to have thoughts and subsequently engage in intellectual and moral agency. We will call this whole package of specifically human capacities *the power of reason.* Though there is much in common between the two, being an actually living human being is not the same thing as being an actually living nonhuman animal, because the nonhuman animal lacks the power of reason.[23] Thus, there must be a distinct soul of the human being that accounts for this difference.

---

deal to be said in favor of Aquinas's theory of knowledge along these lines. See Peifer, *The Concept in Thomism,* for an excellent discussion of how Thomist epistemology can avoid skeptical problems; John Haldane, "A Return of Form in the Philosophy of Mind," and "Insight, Inference, and Intellection," *Proceedings of the American Catholic Philosophical Association,* 73 (2006): 31–45; Stump, *Aquinas,* 217–43; and Fergus Kerr, *After Aquinas: Versions of Thomism* (Oxford: Blackwell, 2002), 17–34. Each of these sources does much to recommend Thomist epistemology to contemporary philosophers. For examples of contemporary philosophers who seem to embrace a quasi-Thomistic position in these matters (even if not explicitly), see Hilary Putnam, *The Threefold Cord: Mind, Body, and World* (New York: Columbia University Press, 1999), 1–70; John McDowel, *Mind and World* (Cambridge, Mass.: Harvard University Press, 1996), 3–24, 66–86; and Wilfrid Sellars, "Being and Being Known," 41–59.

22. John Haldane discusses the sense in which Aquinas can be said to be a naturalist about sensation in "Insight, Inference, and Intellection," 37–39. See also Pouivet, *After Wittgenstein, St. Thomas,* 108–11. Both Aristotle and St. Thomas believe that their philosophy of nature ultimately implies theism, so they can be said to be "naturalists," even about nonhuman organisms and the nonliving world, only in a very limited sense.

23. Of course there is some controversy on this point, but we do not have the luxury of wading through the vast scientific and philosophical literature that surrounds it. For a defense of my assumption that human beings have a power of reason utterly lacked by

William's soul just is whatever accounts for William's being an actual being with the power of reason, instead of some other kind of thing matter might have composed. The relation between the human soul and the human body, like the plant soul–plant body and the animal soul–animal body relation, is an instance of the general form-matter relation we discovered in the philosophy of nature. Neither the human soul nor the human body is a *substance* in itself. They are constituents of a single substance; as Aquinas puts it, "Soul and body are not two actually existent substances. Rather, from the two come one actually existing substance."[24] Moreover, the human soul is inclusive of the power to have thoughts and reason along with the powers requisite for life and consciousness. Thus, being a conscious animal is not some accidental feature of the human being. The human being is an organism of a certain type, a living body; the fact that we have souls irreducibly different from those of lower animals does not exempt us from being animals of a very specific type.[25] The Aristotelian is happy to point out that, even though the power of reason cannot be reduced to the power of sensation, human beings cannot, under natural circumstances, exercise their power of reason without a corresponding act of sensation. That is, thought must be *about something*, it must have an object, so William's thinking presupposes the images, sensations, emotions, and so on, drawn from his conscious experiences. Thus, our power of reason has our neurophysiologically emergent power of sensation as a necessary condition for its exercise under natural conditions. If we are unable to have sensations, we are unable to reason.[26]

There are, nevertheless, two extremely important features of

animals, see Machuga, *In Defense of the Soul;* Lynn Rudder Baker, *Persons and Bodies: A Constitution View* (Cambridge: Cambridge University Press, 2000), 62–64; and John Haldane, *Reasonable Faith* (New York: Routledge, 2010), 120–28.

24. Aquinas, *Summa contra Gentiles,* bk. I, ch. 69.

25. This view is often referred as *animalism* about human nature. For a general defense, see Patrick Toner, "Hylomorphic Animalism," *Philosophical Studies* 155 (2011): 65–81; and George and Lee, *Body-Self Dualism in Contemporary Ethics and Politics,* 1–49.

26. Kenny, *Aquinas on Mind,* 41–58, is helpful on this issue.

the human soul that significantly distinguish it from the animal soul; that is, *the human soul is separable from the human body* and subsequently *it cannot have emerged through any physical process.* We discussed in chapter 6 an argument to the effect that thoughts cannot emerge from neurophysiological processes, because they have universal content, which contrasts with the account of sensation we discussed in the previous section of this chapter. Our considerations in subsequent chapters will help illuminate this argument. Consider the following remarks from St. Thomas:

> The things which belong to the *species* of a material thing, such as a stone, or a man, or a horse, can be thought of apart from the *individualizing* principles which do not belong to the notion of the *species*. This is what we mean by abstracting the universal from the particular, or the intelligible *species* from the phantasm; that is, by considering the *nature* of the *species* apart from its *individual qualities* represented by the phantasms.[27]

Aquinas is availing himself of a great deal of scholastic terminology in this passage, but the point he is making is merely that, whereas a sensation is formally identical with its object as an individual, in thought the object is taken in abstraction from its individuating characteristics, as universal. That is, when I have a sensation of red I become aware of this particular instance of red (an individual red object), at this time, this place, with this shape, and so on, whereas in thought I am able to strip the individuating features of the object, and consider it solely in terms of a universal that can be applied to diverse objects with different individuating features. As John Haldane puts it, "the immateriality of thought is implied by the fact that its objects … are abstract. In perception the sensible form is 'received' into the sense without the original matter of its source, but nonetheless under material conditions of the sense organ," but when we think of the universal natures of things "these various features are entertained as purely abstract, and the content of my thinking does not reflect

27. *Summa theologica,* Pt. I, Q. 85, A. 1.

the thought-organ."[28] There is nothing metaphysically otiose or deeply implausible about claiming that a physical substance has the potency to become red or hot, and so on, so there is nothing blocking an emergentist account of sensation. Being universal is not part of the natural potency of any physical substance, so the coming to be as universal in thought is not something that can emerge from a physical substance. In short, thoughts are not emergent features of neurophysiological processes. Thus, though the power of reason has neurophysiologically emergent processes as a necessary condition for their exercise under natural conditions, such processes are insufficient to account for the universal content of human thoughts (which is an integral part of the power of reason). Therefore, the human soul acts independently of any bodily organ inasmuch as exercising the power of reason requires thoughts with universal content. Remember our dictum that *activity follows on being,* which leads Aquinas to argue as follows:

> Therefore the intellectual principle itself, which is called the mind or the intellect, has an operation *per se,* in which it does not communicate with the body. But nothing can operate *per se,* unless it subsists *per se.* For any entity does not operate unless it is in act: hence a thing operates in the manner in which it exists.[29]

In other words, the human soul, part of which is the intellect or mind, has an activity that simply cannot be an activity of any particular bodily organ. It then satisfies the criteria for Aristotle's exception to his general rule that a soul cannot exist without its body.

It might be helpful to lay this argument out a bit more clearly:

(1) The human soul is the principle that accounts for the power of reason in the human organism. (premise)

(2) The power of reason is exercised without a bodily organ. (premise)

28. Haldane, *Reasonable Faith,* 123.
29. *Summa theologica,* Pt. I, A. 75, A. 2.

(3) If a power can be exercised without a bodily organ, then the principle of that exercise can exist without a body. (premise)

Therefore:

(4) The principle of the power of reason is exercised without a bodily organ. (from 2 and 3)

Therefore:

(5) The human soul can exist without a body. (from 1 and 4)

Premise (1) is a definition of "human soul," and (4) and (5) follow validly from other premises. Premise (2) is the result of an argument we discussed in chapter 6 in great detail and in the previous paragraphs. Briefly, since thoughts have universal content, they cannot be emergent features of any physical system. Of course, thought requires acts of sensation from which universals can be abstracted, so in that sense they have physical processes as necessary conditions. There is not, however, any organ of thought from which thoughts emerge, because of their universal content. Whatever does the work of abstraction is not a bodily organ.

Premise (3) is the crux of the argument, but it is an application of our principle that act follows on being; that is, we posit separable existence based on distinct activities. If a certain power is such that it is beyond the capacity of entities of a certain type to exercise it, then exercise of this power implies that a distinct type of thing exists—in order to *act* in a certain manner, something must *be* in an appropriate manner. We have, given (2), good reason to conclude that thoughts cannot be caused by physical processes (though physical processes are necessary preconditions for thoughts), so whatever causes a thought has a nonphysical activity. If we were to suppose that this principle of thought were itself something physical or dependent on something physical, we would run afoul of what we concluded in (2). Something has to do the work of abstraction in thought; since no bodily organ can do this work, we should then conclude that the principle of thought is something independent of any bodily organ for its

existence. As Patrick Lee and Robert George put it: "The character of an action reveals the character of its principle or source. So if there are actions that are independent of matter (in the sense of being performed without a material organ), then the source of these actions must in some way be or exist independently of matter."[30] The point of the argument is that since a human being can do something in virtue of his soul that is not identical to, supervenient on, or even emergent from its physical constituents, it follows that his soul has an independent existence: "unlike the senses, the intellect has no bodily organ. For the nobility of the human soul transcends the scope and limits of bodily nature. Hence, it enjoys a certain activity in which matter has no share; the activity is without a bodily organ; and in this sense it is a 'separate' intellect."[31] Even though the soul is not a substance in itself, because without the body it cannot naturally exercise its powers, the soul is a *subsistent entity* in the sense that it can simply exist independently of the body. Aquinas goes on to draw an even stronger conclusion: not only is the rational soul a subsistent entity, it is incorruptible (immortal), because it is absolutely simple. Corruptible things are composites of matter and form, but the rational soul is a substantial form that itself cannot be a composite of matter and form. We posit forms to make sense of the unity of physical things (to avoid the problem of the many), but if forms themselves needed *unifiers,* we would be on our way to an infinite regress. Thus, the human soul is not only subsistent, but also incorruptible.[32]

It is important to note that it is not the *human person,* which is a certain kind of organism, but the *human soul* that is separable from the body and incorruptible. William is a human being, and as such he is the sort of organism that, under standard conditions,

30. George and Lee, *Body-Self Dualism,* 67.

31. Aquinas, *Commentary on Aristotle's De Anima,* lecture 7, §699.

32. See *Summa theologica,* Pt. I, Q. 75, A. 6. See Kenny, *Aquinas on Mind,* 129–44, for a critical discussion of Aquinas on the separability and incorruptibility of the human soul.

develops and exercises the power of reason. He is not merely the thing that thinks, however, but also the thing that feels, eats, grows, and so on. In other words, because William's exercise of his power of reason has his animal nature as a necessary precondition, William taken as the rational being possessing the power of reason is one and the same as William the animal organism possessing the power of sensation. Aquinas makes this point clear: "sensation is not the operation of the soul only. Since, then, sensation is an operation of man . . . it is clear that man is not a soul only, but something composed of soul and body."[33] One way to put it is that for Aquinas, it is not William who will survive his bodily death immediately, but William's soul.[34] William is an organism with the power of reason, and something that is not an organism is not William. Indeed, without a body William's soul cannot even think (except possibly through a special act by God), because it would be without the sensations and images that are necessary preconditions for human thinking. Thus, even though his soul is incorruptible, if *William,* not just *William's soul,* is to survive death, there must be a resurrection of *William's body.* This is why Aristotelians such as Aquinas take the plausibility of a human afterlife to hang on the plausibility of the resurrection of the body.[35]

As I mentioned earlier, the human soul is not only separable

33. *Summa theologica,* Pt. I, Q. 75, A. 4.

34. This is actually a controversial point among followers of Aquinas. Some Thomists argue that William will survive the death of his body even prior to his bodily resurrection; see Stump, *Aquinas,* 51–54; and "Resurrection, Reassembly, and Reconstitution: Aquinas on the Soul," in *Die menschliche Seele: Brauchen wir den Dualismus?* ed. Bruno Niederberger and Edmund Runggaldier, 151–72 (Frankfort: Ontos Verlag, 2006). Patrick Toner criticizes this "survivalist" position in "St. Thomas on Death and the Separated Soul," *Pacific Philosophical Quarterly* 91 (2010): 587–99. See also Patrick Toner, "Thomas Versus Tibbles: A Critical Study of *Brown's Aquinas and the Ship of Theseus," American Catholic Philosophical Quarterly* 81, no. 4: 639–53, and Christopher M. Brown, "Souls, Ships, and Substances: A Reply to Toner," *American Catholic Philosophical Quarterly* 81, no. 4 (2007): 655–68. See also John Haldane, "The Examined Death," in *Reasonable Faith,* 143–59.

35. For treatments of the possibility of bodily resurrection by Aristotelian hylomorphists, see the selection for St. Thomas's *Commentary on First Corinthians,* in *Aquinas: Selected Philosophical Writings,* trans. Timothy McDermott (Oxford: Oxford University Press, 1993), 192–93; Geach, *God and the Soul,* 17–29; and George and Lee, *Body-Soul Dualism in Contemporary Ethics and Politics,* 74–81.

(and indeed incorruptible), but it cannot have its origin, that is, it cannot have come to be in the first place, through any physical process. Just as activity follows on being, *coming to be follows upon being.* If the soul has a power that is exercised independently of any physical process and is therefore separable from any such process, it would stand to reason that the soul cannot have originally come to be, in either the individual or the species, through a physical process. Since the soul exists in a nonphysical way, it is difficult to conceive how it could come to be by a physical process.[36] We might, nevertheless, entertain a broader sense in which the human soul could be said to be emergent. That is, the human soul could be said to be *predictively* or *nomologically emergent:* a certain organization of a certain kind of constituents, for example, human DNA structuring an organism, is necessary and sufficient condition for us *to conclude* that the resulting being is the kind of thing capable of human thought and agency, and therefore has a human soul, even though these constituents cannot be the ultimate cause of the soul coming to be. Thus, although the hylomorphist argues that we cannot say that any physical structure explains the power of rationality, he need not deny that empirical investigation can, and increasingly does, discover the physical conditions in which it arises.[37]

This argument will immediately raise eyebrows because of its implications for the possibility of human evolution by natural selection, and indeed emergent dualists are often motivated by the fact that the human species has its origin in the same evolutionary processes as formed all other contemporary species of animals. Simply put, *if the foregoing account of the human soul is correct, evo-*

36. See George and Lee, *Body-Self Dualism in Contemporary Ethics and Politics,* 71, where they apply this Thomistic insight directly to Hasker's emergent dualism. See also Aquinas, *Summa theologica,* Pt. I, Q. 75.

37. Brian Leftow articulates a similar view in "Souls Dipped in Dust," in *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons,* ed. Kevin Corcoran, 120–38 (Ithaca, N.Y.: Cornell University Press, 2001) and "Soul, Mind, and Brain," in *The Waning of Materialism,* ed. Robert C. Koons and George Beeler, 395–417 (New York: Oxford University Press, 2010).

lution by natural selection, or any other entirely natural story, cannot be an explanation of the origin of the human soul. The human soul cannot have evolved in the human species, because it is not a physical thing dependent on physical processes. Moreover, the human soul is metaphysically simple, so it cannot have come to be through an incremental process; a human soul is an all-or-nothing proposition. For somebody who takes the position I am presenting in this chapter, there is no way around the fact that the human soul is not a product of natural evolution. However, there is absolutely nothing in this position that precludes that our status as living and conscious animals is a product of an evolutionary history that we have in common with all living things, and our bodily lives are not merely accidental to us. As best we can tell, our powers of nutrition, growth, reproduction, and consciousness, all of which are necessary conditions for the power of reason, are derived from similar powers among nonhuman living beings through the process of natural selection. Moreover, given that we are organisms, it is not surprising that many of our social practices can be explained as part of a continuum of such behaviors reaching back into lower animals.[38] Thus, *when I assert that the human soul has not evolved, I do not claim that there is some empirical gap that we expect to find in natural history.* At some point, through a nonnatural source, the human soul came to be (and with it the human being), but this supernatural event likely occurred only after a long preparatory process of animal evolution. The origin of the human soul is not empirically detectable, and what we can say of it is only what we can demonstrate philosophically. Thus, nothing that the Aristotelian has to say about the origin of the human soul should make the least bit of difference for the actual practice of evolutionary biology. It is not surprising that we have much to learn about ourselves, our cultures, and our relationship to the rest of the living

38. See MacIntyre, *Dependent Rational Animals,* for a discussion of the continuity between human and nonhuman animal social practices by an Aristotelian hylomorphist.

world from the insights of modern evolutionary biology, even if the human soul is separable, incorruptible, and the product of supernatural origins.[39]

But Is This Dualism?

In chapter 3, I highlighted four problems that hamper dualism: (a) the problem of mind-body interaction; (b) the problem of psycho-physical dependence; (c) the problem of consciousness in nonhuman animals; and (d) the moral implications of denying our psycho-physical unity. I maintain that none of these difficulties is an absolute "deal-breaker" for the dualist, though in dealing with them she must give increasingly difficult philosophical accounts that harm the overall plausibility of her position. It would be quite advantageous if we could avoid these problems altogether without resorting to any version of naturalism, which itself raises as many problems as it solves. It is also quite common for those who are not well acquainted with the position to lump Aristotelian hylomorphism into the same category as dualism and then to reject it because it is subject to some of these problems endemic to dualism.[40] I want to address this mistaken conception of Aristotelian hylomorphism and recommend it at the expense of dualism by showing why none of the typical problems of dualism arises for the hylomorphist. I close this section by discussing why Aristotelian hylomorphism is not plagued by the difficulties regarding intellectual and moral agency that un-

39. For broadly compatibilist accounts of hylomorphism and evolutionary biology see Reichmann, *Evolution, Animal "Rights," and the Environment,* 13–108; Gilson, *From Aristotle to Darwin and Back Again;* and Bobik, in *Aquinas on Matter, Form, and the Elements,* 219–28. See also Christopher Cardinal Schornborn, *Chance or Purpose: Creation, Evolution, and a Rational Faith* (San Francisco: Ignatius Press, 2007), 105–26.

40. See Michael Tooley's contribution to Michael Tooley, Celia Wolf-Devine, Philip Devine, and Alison M. Jaggar, *Abortion: Three Perspectives* (Oxford: Oxford University Press, 2009), 15–20, for an example of a rejection of the hylomorphic view along with dualism in light of the fact of psycho-physical dependence. See also Hud Hudson, "I Am Not an Animal," in *Persons: Human and Divine,* ed. Peter van Inwagen and Dean Zimmerman, 216–34 (New York: Oxford University Press, 2007); Eric Olsen, *What Are We? A Study in Personal Ontology* (Oxford: Oxford University Press, 2007), 171–75.

dermine naturalism in both its materialist and emergentist forms.

(a) *The problem of mind-body interaction.* Take the mind-body interaction problem, in any of its various guises, for example, supposed unintelligibility, conservation laws, the causal closure of the physical, and the like. Even if the dualist can answer these worries (and at the end of the day he might be able to do so), such a case needs to be made. On the dualist model, we have two fundamentally different substances, one physical and the other nonphysical, interacting as efficient causes, even though the laws of physics determine the physical substance. This problem does not even arise for the Aristotelian hylomorphist, however, because the fundamental relationship between the soul and the matter composing the body is not that which might adhere between two independently subsistent entities. That is, the soul and the body are not fundamentally different kinds of *substances* engaged in causal interaction, because they aren't substances at all. Of course, the human soul is a subsistent entity, but the prime matter that ultimately composes the human body has no actuality independent of some substantial form or other. The point is that on the Aristotelian hylomorphist understanding, the body is not a collection of discrete particles that maintain their own substantial identity when entering into the bodily composition. The matter composing a human being is actually a human body only because it is in union with the soul, and the physical particles composing the body are only virtually present within the organism. Thus, it is not as though the soul and the body "push and pull" on each other just like two distinct objects. The actuality of the body is had only through the soul. The issue of interaction arises only if we assume we are dealing with two discrete substances or subsistent entities, which we are not. Of course the doctrine of virtual presence is not something that contemporary philosophers are lining up to defend, but it is a consequence of Aristotelian hylomorphism, and it does avoid the mind-body interaction problem entirely.

(b) *The problem of psycho-physical dependence.* It is quite obvi-

ous that the mind is dependent on the brain in a profound way, and this is not what we would expect if dualism were true. The dualist will often reply that we can account for this fact by arguing that, for all we know, the mind of a brain-damaged person is functioning, though he is getting faulty information from the brain. This sort of story is, of course, possible, but it is adding to the increasingly far-fetched story the dualist is winding. Sensations, on the Aristotelian hylomorphist view I am defending, are emergent from neurophysiological processes, and therefore the fact that such psychological states depend on the proper functioning of the brain is exactly what I would expect to find. Moreover, psycho-physical dependence is also what we would expect to find in the case of thoughts, given that the intellectual part of the human soul requires particulars on which to perform abstraction. The intellect, no less than our faculty of sensation, cannot function in the least without a properly functioning nervous system, at least prior to bodily death. Thus, specifically human, rational consciousness is physically dependent on (though not reducible to) physical processes in the nervous system. Suffice it to say, it would not be new information to Aristotle if a twenty-first-century naturalist philosopher were to point out to him that head injuries, drunkenness, and anomalies in our physiology can render us unable to reason. Of course, in all fairness, this would not be shocking news to Descartes either, but given the dualist story, we wouldn't expect psycho-physical dependence. It is, however, what we would expect given hylomorphism in general.

(c) *The problem of consciousness in nonhuman animals.* It is quite obvious that nonhuman animals have sensations, which would seem to imply that we must give a similar account of their consciousness as we do human consciousness. A substance dualist would have to attribute nonphysical, separable souls to all conscious animals, or at least those that have higher levels of consciousness. The Aristotelian hylomorphist can avoid positing so many separable souls without denying the fact of conscious-

ness in nonhuman animals by granting (at least the possibility of) an emergence or property dualist account of sensations (which is plausible for a hylomorphic philosophy of nature, but not for mechanism), while maintaining that there is a distinction in principle between sensations and distinctively human thoughts based on the universal content of the latter and the particularity of the former. The emergence of sensation is perfectly consistent with the principles of Aristotelian philosophy of nature, so there is no need to deny that nonhuman animals have sensations. Thus, hylomorphism allows us to admit the continuity of human and nonhuman animal powers of sensation, without needing to deny the discontinuity between human and nonhuman animals with respect to the power of reason and the consequent separability of the human soul.

(d) *The moral implications of conceiving of ourselves as souls occupying bodies.* Many critics of dualism are concerned with certain moral implications of conceiving of persons as souls or minds that occupy human bodies, rather than as organisms. The worry is that if we take William's body as something external to what he is essentially, then in harming the body, we are not really harming William, but only something to which he is related instrumentally; William's body is not him, but a tool he uses to pursue his ends. Of course it would be wrong to destroy or steal William's rightful property without his consent, so even on a dualist understanding of the person-body relationship it might be morally wrong in some cases to destroy a human body. The problem is that murder isn't merely a species of theft, and relegating the body to an instrument of the mind cannot help but to have a cheapening effect on the value of human bodily existence. The *Aristotelian hylomorphist can avoid this problem simply because he conceives of the human person as nothing more or less than a human organism.* William is a living human body. Certainly, William has a power of reason that transcends the powers of any of his physical organs, but even that power is rendered inert when separated

from William's life as a human organism. The Aristotelian hylomorphist takes both categories mentioned in the traditional definition of "man" as a *rational animal* perfectly seriously. *Human existence, though it ultimately transcends the physical, is a form of animal existence.*[41]

(e) Finally, *Aristotelian hylomorphism can accommodate both intellectual and moral agency.* Naturalism, in any of the renderings we have discussed, must take thought and action as entirely determined by the microphysical particles that compose human bodies. In our previous discussion, the main problem facing rational agency for naturalists is the impossibility of giving an account of thoughts as emergent from neurophysiological elements, because of the universal content in thinking. The difficulty, as we saw earlier, is that it is impossible to account for universal content as emergent from particular, physical components of a system, and without universal content it is impossible to explain an agent's beliefs in terms of logically valid reasoning. Since the Aristotelian argues that the human intellectual power is not emergent from a neurophysiological system and therefore nothing bars an account of the universal content of thought, this difficulty does not arise. Human beings can give reasons for and justify their beliefs, because we are able to grasp universals that ground our inferences.

The problem of moral agency arises because the naturalist must claim that our actions are accounted for by prior physical conditions (including prior neurophysiological states), which excludes explanations of our actions in terms of goods we choose to pursue. We have seen throughout our discussion in the last three chapters, however, the Aristotelian hylomorphist considers the power of human reason as something that transcends complete physical explanation, and this is likewise the key to moral

41. See George and Lee, *Body-Self Dualism in Contemporary Ethics and Politics,* for a discussion of the moral implications of the hylomorphist theory of human nature.

agency. Thus, even though our desires and inclinations may be determined physically, "these inclinations are subject to the judgment of reason, which the lower appetite obeys."[42] That is, even though our desires and inclinations may not be something for which we can give reasons, there is no threat to moral agency:

> But on the part of the body and its powers man may be such by virtue of a natural quality, inasmuch as he is of such a temperament or disposition due to any impression whatever produced by corporeal causes, which cannot affect the intellectual part, since it is not the act of a corporeal organ. And such as a man is by virtue of a corporeal quality, such also does his end seem to him, because from such a disposition a man is inclined to choose or reject something. But these inclinations are subject to the judgment of reason, which the lower appetite obeys, as we have said [Pt. I, Q. 81, A. 3]. Wherefore this is in no way prejudicial to free-will.[43]

In other words, though the Aristotelian hylomorphist does not deny that human beings are embodied beings, subject to physiologically based desires and inclinations, the fact that human beings have a transcendent intellectual power that operates independently of any bodily organ in its proper activity undermines the concern that we are merely subjects of our environment and physiology. Inasmuch as our acting on our desires and inclinations can be influenced by our intellectual power, there is no reason to question our status as moral agents. In other words, because the human intellect transcends any physical determinism and can grasp universal content, we are able to give reasons that significantly explain our actions in terms of goods we choose to pursue. Moreover, because the human soul is metaphysically more fundamental than the virtually present physical constituents of the human body (the latter exist as constituents of a human body only through the actuality of the human soul), there is no reason to suppose that nature is causally closed to the powers of the human soul. Thus, since the human being has powers that

42. *Summa theologica,* Pt. I, Q. 83, A. 1.
43. Ibid., Pt. I, Q. 83, A. 2.

transcend the physical and the physical is receptive to the activity of these powers, there is no reason to doubt that some of our activities are explained not by causes acting upon us, but by our will to pursue certain goods.[44]

A Key Objection and Reply

I want to close our discussion by addressing what I take to be a well-placed, though ultimately answerable, objection to the application of Aristotelian hylomorphism to the philosophy of mind. When Aristotelian hylomorphists (including Aristotle himself) give examples to illuminate their position, they often use claims like "In a statue the form is the shape and the matter is the clay," or "In a desk the form is the configuration of the wood and the wood is the matter." If statues and desks were bona fide physical substances, then indeed their forms would be the shape of clay and the configuration of the wood, because shape and configuration are sufficient to account for the difference between an actual statue or desk and a potential statue or desk. Often we find definitions of form as "the organizational principle" or "the material configuration" of a physical substance. Organizational states and configurations are definitely not subsistent entities, so if the hylomorphist thinks of the soul along these lines, he has definitely distinguished his position from dualism. Likewise, the statue is not identical to the clay, but to the clay *in a certain configuration,* so defining the soul in this way also distinguishes hylomorphism from reductive materialism. The human soul, as a form, is then just "an essentially configurational state."[45]

When Aristotelians, following Aquinas, argue that the human soul is a subsistent entity, many critics believe that they have

44. Helpful discussions of St. Thomas on moral agency can be found in Kenny, *Aquinas on Mind,* 74–88, and Pouivet, *After Wittgenstein, St. Thomas,* 83–97.

45. Eleonore Stump, "Non-Cartesian Substance Dualism and Materialism without Reductionism," *Faith and Philosophy* 12, no. 4 (1995), 519. Edward Feser also defines form as an "organizational structure" when defending a version of hylomorphism in *Philosophy of Mind,* 220.

contradicted the definition of the soul as a form. Kevin Corcoran presses the hylomorphist on an ambiguity in the use of the word "form."

> Aquinas's view of the soul, however, is not without difficulties. Some of the difficulties center on Aquinas's ambiguous use of the term soul, using it sometimes in the sense of a particular thing and at other times in the sense of "form" or kind of "state" a body is in. Perhaps most important, however, is how, given the rest of Aquinas's metaphysics of substance, a soul can plausibly be said to survive the death of the body.[46]

The problem, and Corcoran is correct to point this out, is that organizational states, configurations, shapes, and the like, are all abstract entities, properties, or ways of being, but definitely not things that can exist without something organized, configured, or shaped. In short, by what seems to be the hylomorphist's own definition of "form," the notion of a separable form is incoherent. According to this interpretation, forms are states of matter, not subsistent entities. As soon as the hylomorphist argues for the separability of the soul, forms seem to cease to be properties or ways of being (as one critic puts it, they "lose their adverbial status"),[47] which contradicts the main thrust of hylomorphist philosophy of nature. Aristotelian hylomorphism is "ambiguous between signifying form as abstract or as concrete."[48] Anthony Kenny offers a criticism of St. Thomas directly using similar terminology, when he claims that Aquinas first

> argues to the conclusion that the soul is incorporeal in the sense that it is abstract and not concrete: it is not a body but the actuality of a body. [Aquinas then argues] to the conclusion that the soul is incorporeal in the sense that it is a non-physical part of a human being: it is an agent with no bodily organ. But an agent cannot be an abstraction, and what is abstract cannot be a part of what is concrete.[49]

46. Corcoran, "The Constitution View of Persons," 162.

47. Gordon Barnes, "The Paradoxes of Hylomorphism," *Review of Metaphysics* 56 (March 2003): 501–23.

48. Hasker, *The Emergent Self*, 166. Hasker also presses this line of criticism in "On Behalf of Emergent Dualism," in *In Search of the Soul*, 94–95.

49. Kenny, *Aquinas on Mind*, 136.

All of these criticisms amount to the same objection: on one hand, the hylomorphist emphasizes form as an abstract entity when distinguishing his position from dualism, and on the other hand, the hylomorphist presents form as though it were a kind of concrete particular (a subsistent entity) when giving an account of human nature and the possibility for survival of bodily death. This all looks like a logically inconsistent case of trying to have your cake and eat it too!

Given the way hylomorphism is presented even by many of its advocates, I agree with this line of criticism; some presentations of hylomorphism do seem to trade on a convenient ambiguity between *form as abstract* and *form as a particular* (*concrete*). Careful attention to what I have said earlier, however, should suffice to dissuade one from understanding hylomorphism in this way. All along, I have taken form and matter as *functional concepts.* Jeffrey Brower supports this rendering of hylomorphism, at least as an interpretation of Aquinas, when he argues that for Aquinas

> the notions of matter, form, and compound … are (at least in the first instance) *purely functional in nature.* To be matter, on this account, is just to be an entity playing a certain function or role—that of accounting for the sameness involved in change (namely, sameness of subject over time). Likewise, to be form, on this account, is to be an entity playing a distinct function or role—that of accounting for the difference involved in change (namely, difference of characterization over time). Finally, to be a hylomorphic compound, on this account, is just to be an entity that possesses such functional matter and form, and hence one capable of being characterized by the special type of coming-to-be and passing-away associated with change (namely, generation and corruption).[50]

In other words, *matter, form, and the matter-form compound cannot be defined except in terms of roles played in change.* The matter of a substance is whatever potency preexists and survives its generation (and will survive its corruption); form is whatever accounts for the difference between the matter and the actual

50. Jeffrey Brower, "Matter, Form and Individuation," in *The Oxford Handbook to Aquinas,* ed. Brian Davies and Eleonore Stump (New York: Oxford University Press, 2012), 88.

substance whose generation or corruption we set out to explain, and the compound substance is whatever individual is subject to this process of change. There is nothing in the general notions of matter and form as they are introduced in the account of change that entails any commitment regarding their status as abstract or concrete entities. As we saw in the previous chapter, Aquinas and Aristotle both introduce matter and form initially as the principles of change, that is, that which accounts for the possibility of change.[51] They are silent as to what these principles are beyond their essential roles in providing the possibility of change.

Anybody who has spent a moment's time in a classroom attempting to teach undergraduate students about the matter-form distinction knows (likely all too well) that it is easiest to bring the distinction to light using examples of artifacts such as statues and tables, and this is a pedagogical technique modeled by Aristotle and Aquinas themselves. In whatever loose sense (if any) statues and tables are substances, their forms are configurations or organizational states. In the case of artifacts, abstract entities of this sort can serve as forms; that is, all that distinguishes the statue from the clay is its shape, and all that distinguishes the table from the wood is the configuration. That is not to say, however, that only an abstract entity can play the functional role of form. If cases arise in which the role of form can be satisfied only by a concrete particular, then in those cases the form will be a subsistent entity. I fear the centuries of the pedagogical demands of introducing the matter-form distinction have clouded this essential part of the hylomorphic doctrine.

There is then nothing ambiguous or incoherent about Aquinas's treatment of forms as organizational states and subsistent entities in different contexts. When dealing with entities whose actuality can be provided for by an organizational state alone,

51. Although he presses Aquinas on the supposed concrete-abstract ambiguity, Kenny actually seems to cast the notions of "matter" and "form" in the sort of functional terms I am recommending. See Kenny, *Aquinas on Mind,* 23–29.

there is no need to posit anything more for its substantial form. The human being, however, is such that its actuality cannot be accounted for by an organizational state alone, because its soul has a power that transcends embodiment. Aristotle famously chides Plato for his claim that abstract objects are the ultimate causes of concrete things, because abstract objects cannot make something happen.[52] In this same way, we should likewise dismiss the claim that the *act* of thinking can be explained by an abstract entity. Abstraction and agency are activities, and activities require concrete particulars to perform them. The power of reason (the capacity for thinking and agency) therefore requires a particular, concrete thing that can exercise such a power, and it is the power of reason that differentiates the actual human being from something not actually (or even potentially) a human being. Thus, the human substantial form (the human soul) must be a concrete particular, not an abstract entity. Given that form is a functional concept, this is all perfectly consistent with the general principles of hylomorphism.

Some Tentative Conclusions

In this and the previous chapter, I have attempted to sketch an Aristotelian-hylomorphic philosophy of nature and draw out its consequences for the philosophy of mind. I have made every effort to highlight those claims inherent to this view that are likely to be counterintuitive, if not downright shocking, to many contemporary philosophers. At the same time, I hope that I have impressed upon you that the hylomorphist embraces these positions only because these are the consequences that come part and parcel of giving a coherent account of change. Some of these positions in the philosophy of nature actually bear great fruits in the philosophy of mind; for example, the doctrines of prime matter and the virtual presence of the elements render the mind-

52. See Aristotle, *Metaphysics,* 191b1–20.

body interaction problem a "nonstarter." When applied to the problems of the philosophy of mind, Aristotelian hylomorphism avoids both the excesses of substance dualism and the needless austerity of naturalism. Please note that hylomorphism is not a wishy-washy attempt at splitting the difference between dualism and naturalism, but a systematic view of nature, inclusive of human nature. It should be accepted or rejected based not solely on its results in the philosophy of mind, but on its merits as a comprehensive philosophy. Of course, I have fallen far short of giving compelling support for anything so broad, so all I can do here is to offer hylomorphism as a proposal bolstered by its strengths in the philosophy of mind (and the weaknesses of its competitors) with some indications of its strengths as a philosophy of nature. Hopefully I have made this proposal a live option even to those of you who have pre-philosophical intuitions that point in other directions.

At any rate, it should be clear to you at this point that there is nothing in the arguments among the philosophers of mind that gives us reason to doubt seriously our status as beings possessing the powers of sensation, thinking, and intellectual and moral agency. There are coherent and scientifically informed ways of approaching mind and nature that paint a picture of humanity as possessing the dignity of freedom and an eternal destiny. This is not to say that there are no looming philosophical problems to be addressed, and none of these issues are settled in the sense that every smart person who thinks about them will come immediately to the same conclusions. This need not bother us, however, once we realize that the point of philosophy is not consensus, but truth, and the fact of disagreement, even among smart people, is never enough to reject a philosophical position for which you have compelling reasons. As long as some of us have the leisure to spend our lives philosophizing, there will always be controversy surrounding human nature. Be that as it may, at some point you will need to come to a conclusion as to which position in

the philosophy of mind is intellectually superior—it would be a shame to spend your entire life without coming to any conclusion as to what kind of being you are. If I have done my job well, this book will have advanced you toward that end, even if you do not end up agreeing with me.

BIBLIOGRAPHY

Adler, Mortimer J. *The Difference of Man and the Difference It Makes.* New York: Fordham University Press, 1993.

———. *Intellect: Mind over Matter.* New York: Macmillan Publishing, 1990.

Almog, Joseph. *What Am I? Descartes and the Mind-Body Problem.* New York: Oxford University Press, 2005.

Anscombe, Elizabeth. "Causation and Determination." In *Causation,* edited by Ernest Sosa and Michael Tooley, 88–104. New York: Oxford University Press, 1993.

———. *From Parmenides to Wittgenstein: Collected Philosophical Papers.* Vol. 1. Minneapolis: University of Minnesota Press, 1981.

———. "The Immortality of the Soul." In *Faith in a Hard Ground: Essays on Religion, Philosophy, and Ethics by Elizabeth Anscombe,* edited by Mary Geach and Luke Gormally, 69–83. Exeter, UK: Imprint Academic, 2008.

Aquinas, St. Thomas. *Aquinas on Matter, Form, and the Elements: A Translation and Interpretation of the De Principiis Naturae and the De Mixtione Elementorum of St. Thomas Aquinas.* Edited and translated by Joseph Bobik with commentary. Notre Dame, Ind.: University of Notre Dame Press, 1998.

———. *Aquinas: Selected Philosophical Writings.* Translated by Timothy McDermott. Oxford: Oxford University Press, 1993.

———. *Commentary on Aristotle's De Anima.* Translated by Kenelm Foster, O.P., and Sylvester Humphries. Notre Dame, Ind.: Dumb Ox Books, 1954.

———. *Commentary on Aristotle's Physics.* Translated by Richard J. Blackwell, Richard J. Spath, and W. Edmund Thirlkel. Notre Dame, Ind.: Dumb Ox Books, 1999.

———. *On Truth.* Translated by Robert W. Mulligan, S.J. Chicago: Henry Regnery, 1952.

———. *Summa contra Gentiles.* Translated by James F. Anderson. Notre Dame, Ind.: University of Notre Dame Press, 1956.

———. *Summa theologica.* Translated by the Fathers of the English Dominican Province. New York: Benziger Bros., 1947.

Aristotle. *The Complete Works of Aristotle: The Revised Oxford Translation.* Edited by Jonathan Barnes. New York: Oxford University Press, 1984.

Armstrong, David. *A Materialist Theory of Mind.* London: Routledge and Kegan Paul, 1968.

———. *The Mind-Body Problem: An Opinionated Introduction.* Boulder, Colo.: Westview Press, 1999.

———. "Naturalism, Materialism, and First Philosophy." *Philosophia* 8, no. 2–3 (1978): 261–76.

———. *Universals: An Opinionated Introduction.* Boulder, Colo.: Westview Press, 1989.

Ayer, A. J., ed. *Logical Positivism.* New York: Free Press, 1966.

Baker, Lynn. *Saving Belief: A Critique of Physicalism.* Princeton, N.J.: Princeton University Press, 1987.

Baker, Mark, and Steward Goetz, eds. *The Soul Hypothesis: Investigations into the Existence of the Soul.* New York: Continuum, 2011.

Barnes, Gordon. "The Paradoxes of Hylomorphism." *Review of Metaphysics* 56 (March 2003): 501–23.

Beakely, Brian, and Peter Ludlow, eds. *Philosophy of Mind: Classical Problems/Contemporary Issues.* Boston, Mass.: MIT Press, 2006.

Beauregard, Mario, and Denyse O'Leary. *The Spiritual Brain: A Neuroscientist's Case for the Existence of the Soul.* New York: HarperOne, 2008.

Behe, Michael. *Darwin's Black Box.* New York: Free Press, 1998.

Beilby, James K., ed. *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism.* Ithaca, N.Y.: Cornell University Press, 2002.

Bergman, Michael. "Commonsense Naturalism." In *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism,* edited by James K. Beilby, 61–90. Ithaca, N.Y.: Cornell University Press, 2002.

Black, Max. "The Identity of Indiscernibles." In *Classics of Analytical Metaphysics,* edited by Larry L. Blackman, 281–92. Laham, Md.: University Press of America, 1984.

Blackmore, Susan. *Conversations on Consciousness: What the Best Minds Believe about the Brain, Free Will, and What It Means to Be a Human.* New York: Oxford University Press, 2006.

———. *Dying to Live: Near-Death Experiences.* Buffalo, N.Y.: Prometheus Books, 1993.

Block, Ned. ed. *Readings in Philosophy of Psychology.* Vol. 1. Cambridge, Mass: Harvard University Press, 1980.

———. "Troubles with Functionalism." In *Mind and Cognition: A Reader,* edited by William Lycan, 444–68. Oxford: Routledge, 1990.

Braine, David. *The Human Person: Animal and Spirit.* Notre Dame, Ind.: University of Notre Dame Press, 1992.

Broad, C. D. "Mechanism and Its Alternatives." In *The Philosophy of Mind: Classical and Contemporary Readings,* edited by David Chalmers, 106–15. New York: Oxford University Press, 2002.

———. *Mind and Its Place in Nature.* London: Routledge and Kegan Paul, 1925.

Brower, Jeffrey. "Matter, Form, and Individuation." In *The Oxford Handbook to Aquinas,* edited by Brian Davies and Eleonore Stump, 85–103. New York: Oxford University Press, 2012.

Brown, Christopher M. Brown. "Souls, Ships, and Substances: A Reply to Toner." *American Catholic Philosophical Quarterly* 81, no. 4 (2007): 655–68.

Cantens, Bernardo. "Cognitive Faculties and Evolutionary Naturalism." *Proceedings of the American Catholic Philosophical Association* 80 (2006): 201–8.

Captain, William H., and Daniel D. Merrill, eds. *Art, Mind, and Nature.* Pittsburgh, Pa.: University of Pittsburgh Press, 1967.

Carnap, Rudolf. "Psychology in Physical Language." *Erkenntnis* 3 (1932–33): 107–42. Reprinted in *Logical Positivism,* edited by A. J. Ayer, 165–98. New York: Free Press, 1966.

Chalmers, David. *The Conscious Mind: In Search of a Fundamental Theory.* New York: Oxford University Press, 1996.

———. "Consciousness and Its Place in Nature." In *The Philosophy of Mind: Classical and Contemporary Readings,* edited by David Chalmers, 247–74. New York: Oxford University Press, 2002.

———. "I'm a Conscious Zombie." In Susan Blackmore, *Conversations on Consciousness: What the Best Minds Believe about the Brain, Free Will, and What It Means to Be a Human,* 36–49. New York: Oxford University Press, 2006.

———, ed. *The Philosophy of Mind: Classical and Contemporary Readings.* New York: Oxford University Press, 2002.

Chesterton, G. K. *Orthodoxy.* San Francisco: Ignatius Press, 1995.

Chisholm, Roderick. "On the Simplicity of the Soul." In *Philosophical Perspectives 5: Philosophy of Religion,* edited by James E. Tomberlin, 167–81. Atascadero, Calif.: Ridgeview Publishing, 1991.

———. *Person and Object.* Chicago: Open Court, 1976.

Churchland, Paul. "Eliminative Materialism and Propositional Attitudes." *Journal of Philosophy* 78, no. 2 (1981): 67–90.

———. *Matter and Consciousness.* 2nd ed. Cambridge, Mass.: MIT Press, 1988.

———. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, Mass.: MIT Press, 1989.

Churchland, Paul, and Patricia Churchland. *On the Contrary: Critical Essays 1987–1997.* Cambridge, Mass.: MIT Press, 1998.

Clarke, W. Norris. *The One and the Many: A Contemporary Thomistic Metaphysics.* Notre Dame, Ind.: University of Notre Dame Press, 2001.

Collins, Robin. "The Energy of the Soul." In *The Soul Hypothesis: Investiga-*

tions into the Existence of the Soul, edited by Mark C. Baker and Stewart Goetz, 123–33. London: Continuum, 2011.
Corcoran, Kevin. "The Constitution View of Persons." In *In Search of the Soul: Four Views of the Mind-Body Problem,* edited by Joel B. Green and Stuart L. Palmer, 153–76. Downers Grove, Ill.: InterVarsity Press, 2005.
Cottingham, John. "Cartesian Trialism." *Mind* 94, no. 374 (1985): 218–30.
Crane, Tim. *The Mechanical Theory of Mind.* London: Penguin Books, 1995
Cunningham, Suzanne. *What Is a Mind?* Indianapolis, Ind.: Hackett Publishing, 2000.
Curd, Patricia. *The Legacy of Parmenides: Eleatic Monism and Later Presocratic Thought.* Princeton, N.J.: Princeton University Press, 1998.
Curd, Patricia, and Richard McKirahan, eds. and trans. *The Presocratics Reader.* Indianapolis, Ind.: Hackett Publishing, 1996.
Danto, Arthur. "Naturalism." In *The Encyclopedia of Philosophy,* edited by Paul Edwards, 448–50. New York: Macmillan, 1967.
Davidson, Donald. "Davidson, Donald." In *A Companion to the Philosophy of Mind,* edited by by Samuel Guttenplan, 231–36. Cambridge, Mass.: Blackwell, 1995.
———. "Mental Events." In *Essays on Actions and Events,* 207–25. Oxford: Clarendon Press, 1980.
Davies, Brian. *Thomas Aquinas: Contemporary Philosophical Perspectives.* New York: Oxford University Press, 2002
Dawkins, Richard. *The Blind Watchmaker.* New York: W. W. Norton, 1996.
De Lubac, Henri. The Drama of Atheistic Humanism. 7th ed. San Francisco: Ignatius Press, 1995.
Dennett, Daniel. *Brainstorms: Philosophical Essays on Mind and Psychology.* Cambridge, Mass.: MIT Press, 1981.
———. *Consciousness Explained.* Boston, Mass.: Back Bay Books, 1991.
———. *Darwin's Dangerous Idea: Evolution and the Meanings of Life.* New York: Simon and Schuster, 1995.
———. *The Intentional Stance.* Cambridge, Mass.: MIT Press, 1989.
———. "Quining Qualia." In *Mind and Cognition,* edited by William Lycan, 519–48. Cambridge, Mass.: Blackwell, 1990.
———. "Why the Law of Effect Will Not Go Away." In *Mind and Cognition,* edited by William G. Lycan, 63–73. Cambridge, Mass.: Blackwell, 1990.
Dennett, Daniel, and Alvin Plantinga. *Science and Religion: Are They Compatible?* New York: Oxford University Press, 2011.
Descartes, Rene. *Philosophical Essays and Correspondence.* Edited by Roger Ariew. Indianapolis, Ind.: Hackett Publishing, 2000.
———. *The Philosophical Writings of Descartes.* Vols. 1–3. Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch. New York: Cambridge University Press, 1985 and 1991.

Dicker, Georges. *Descartes: An Analytical and Historical Introduction.* New York: Oxford University Press, 1993.

Dostoevsky, Fyodor. *The Grand Inquisitor.* Translated by Constance Garnett. Indianapolis, Ind.: Hackett Publishing, 1993.

Dretske, Fred. *Naturalizing the Mind.* Cambridge, Mass.: MIT Press, 1995.

Elders, Leo. *The Philosophy of Nature of St. Thomas Aquinas: Nature, Universe, Man.* New York: Peter Lang, 1997.

Epicurus. "Letter to Herodotus." In *The Epicurus Reader.* Edited and translated by Brad Inwood and L. P. Gerson. Indianapolis: Hackett Publishing, 1994.

Feigl, Herbert. "The 'Mental' and the 'Physical.'" In *Concepts, Theories, and the Mind-Body Problem,* edited by H. Feigl, G. Maxwell, and M. Scriven, 370–497. Minnesota Studies in the Philosophy of Science 2. Minneapolis: University of Minnesota Press, 1958.

Felt, James. *On Human Knowing: A Prelude to Metaphysics.* Notre Dame, Ind.: University of Notre Dame Press, 2005.

Feser, Edward. *Philosophy of Mind: A Beginner's Guide.* Oxford: Oneworld Publications, 2005.

Feyerabend, Paul. "Mental Events and the Brain." In *Materialism and the Mind Body Problem,* 2nd ed., edited by David Rosenthal, 172–74. Indianapolis, Ind.: Hackett Publishing, 2000.

Flanagan, Owen. *Consciousness Reconsidered.* Cambridge, Mass.: MIT Press, 1992.

———. *The Problem of the Soul.* New York: Basic Books, 2002.

Flew, Antony. "Death." In *New Essays in Philosophical Theology,* edited by Antony Flew and Alasdair MacIntyre, 267–72. London: SCM Press, 1955.

———. *God and Philosophy.* Amherst, N.Y.: Prometheus Books, 2005.

———. *There Is a God: How the World's Most Notorious Atheist Changed His Mind.* New York: Harper Collins, 2007.

Fodor, Jerry. "Materialism." In *Materialism and the Mind Body Problem,* 2nd ed., edited by David Rosenthal, 128–49. Indianapolis, Ind.: Hackett Publishing, 2000.

———. *A Theory of Content and Other Essays.* Cambridge, Mass.: MIT Press, 1990.

Foster, David Ruel. "Aquinas on the Immateriality of the Intellect." *Thomist* 55 (1991): 415–38.

Foster, John. "A Brief Defense of the Cartesian View." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons,* edited by Kevin Corcoran, 17–18. Ithaca, N.Y.: Cornell University Press, 2001.

———. *The Immaterial Self.* New York: Routledge, 1991.

Freddoso, A. J. "Suarez on Metaphysical Inquiry, Efficient Causality, and Divine Action." In Francisco Suarez, *On Creation, Conservation, and Con-*

currence: Metaphysical Disputations 20–22, translated by A. J. Freddoso, xi–cxxi. South Bend, Ind.: St. Augustine's Press, 2002.
Geach, Peter. *God and the Soul.* South Bend, Ind.: St. Augustine Press, 1969.
———. *Mental Acts.* London: Routledge and Kegan Paul, 1957.
———. *Reference and Generality.* Ithaca, N.Y.: Cornell University Press, 1980.
George, Robert. *The Clash of Orthodoxies: Law, Religion, and Morality in Crisis.* Wilmington, Del.: ISI Books, 2001.
George, Robert, and Patrick Lee. *Body-Self Dualism in Contemporary Ethics and Politics.* New York: Cambridge University Press, 2009.
Gilson, Etienne. *Being and Some Philosophers.* Toronto: Pontifical Institute for Medieval Studies, 1952.
———. *From Aristotle to Darwin and Back Again.* Translated by John Lyon. Notre Dame, Ind.: University of Notre Dame Press, 1984.
Goetz, Stewart. "Modal Dualism: A Critique." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons,* edited by Kevin Corcoran, 89–104. Ithaca, N.Y.: Cornell University Press, 2001.
———. "Naturally Understanding Naturalism." *Faith and Philosophy* 27, no. 1 (2004): 79–90.
———. "Substance Dualism." In *In Search of the Soul: Four Views of the Mind-Body Problem,* edited by Joel B. Green and Stuart L. Palmer, 33–60. Downers Grove, Ill.: InterVarsity Press, 2005.
Goetz, Stewart, and Charles Taliaferro. "Reply to Melnyk's Objections." http://www.infidels.org/library/modern/stewart_goetz/against-melnyk.html.
———. *Naturalism.* Grand Rapids, Mich.: Eerdmans, 2008.
Gutting, Gary. *What Philosophers Know: Case Studies in Recent Analytic Philosophy.* New York: Cambridge University Press, 2009.
Habermas, Gary, and J. P. Moreland. *Beyond Death: Exploring the Evidence for Immortality.* Wheaton, Ill.: Crossway Books, 1998.
Haldane, John. "Analytic Philosophy and the Nature of Mind: Time for Another Rebirth." In *The Mind-Body Problem: A Guide to the Current Debate,* edited by Richard Warner and Tadeusz Szubuka, 197–203. New York: Wiley-Blackwell, 1994.
———. "On Coming Home to (Metaphysical) Realism." *Philosophy* 71 (1996): 287–96.
———. "Insight, Inference, and Intellection." *Proceedings of the American Catholic Philosophical Association* 73 (1999): 31–45.
———. "The Metaphysics of Intellect(ion)." *Proceedings of the American Catholic Philosophical Association* 80 (2006): 39–55
———. *Reasonable Faith.* New York: Routledge, 2010.
———. "A Return to Form in the Philosophy of Mind." In *Form and Matter: Themes in Contemporary Metaphysics,* edited by David Oderberg, 40–64. Oxford: Blackwell, 1999.

Haldane, John, and J. J. C. Smart. *Atheism and Theism.* Malden, Mass.: Wiley-Blackwell, 1996.

Hart, W. D. *The Engines of the Soul.* Cambridge: Cambridge University Press, 1988.

Hasker, William. *The Emergent Self.* Ithaca, N.Y.: Cornell University Press, 1998.

———. "On Behalf of Emergent Dualism." In *In Search of the Soul: Four Views of the Mind-Body Problem*, edited by Joel B. Green and Stuart L. Palmer, 75–100. Downers Grove, Ill.: InterVarsity Press, 2005.

———. "Persons and the Unity of Consciousness." In *The Waning of Materialism,* edited by Robert Koons and George Bealer, 175–90. New York: Oxford University Press, 2010.

———. "Persons as Emergent Substances." In *Body, Soul, and Survival: Essays on the Metaphysics of Human Persons,* edited by Kevin Corcoran, 107–19. Ithaca, N.Y.: Cornell University Press, 2001: 107–19.

Haugeland, John. "What Is Mind Design?" In *Mind Design 2,* edited by J. Haugeland, 1–28. Cambridge, Mass.: MIT Press, 1997.

Heil, John, ed. *The Philosophy of Mind: A Guide and Anthology.* New York: Oxford University Press, 2004.

Hemple, Carl. "The Logical Analysis of Psychology." In *The Philosophy of Mind: A Guide and Anthology,* edited by John Heil, 85–95. New York: Oxford University Press, 2004.

Hill, Christopher. "Imaginability, Conceivability, Possibility, and the Mind-Body Problem." *Philosophical Studies* 87, no. 1 (1997): 61–85.

Horgan, Terence. "From Supervenience to Superdupervenience." In *The Philosophy of Mind: Classical and Contemporary Readings,* edited by David Chalmers, 150–62. New York: Oxford University Press, 2002.

Horgan, Terence, and James Woodward. "Folk Psychology Is Here to Stay." *Philosophical Review* 94, no. 2 (1985): 197–220.

Hudson, Hud. "I Am Not an Animal." In *Persons: Human and Divine,* edited by Peter van Inwagen and Dean Zimmerman, 216–34. New York: Oxford University Press, 2007.

Hume, David. *Dialogues Concerning Natural Religions, with Other Essays.* Indianapolis, Ind.: Hackett Publishing, 1980.

———. *An Enquiry Concerning Human Understanding.* Indianapolis, Ind.: Hackett Publishing, 1977.

———. *Treatise on Human Nature.* New York: Oxford University Press, 1978.

Huxley, Thomas. "On the Hypothesis That Animals Are Automata and Its History." http://alepho.clarku.edu/huxley/CE1/AnAuto.html.

Jackson, Frank. "Epiphenomenal Qualia." *Philosophical Quarterly* 32 (1982): 127–36. Reprinted in *Mind and Cognition,* edited by William G. Lycan, 469–77. Cambridge, Mass.: Blackwell Publishers, 1990.

———. *From Metaphysics to Ethics.* New York: Oxford University Press, 1998.
Jaworski, William. *The Philosophy of Mind: A Comprehensive Introduction.* New York: Blackwell, 2011.
Kant, Immanuel. *The Critique of Pure Reason. Translated by Norman Kemp Smith.* New York: St. Martin's Press, 1965.
Kenny, Anthony. *Aquinas on Mind.* New York: Routledge, 1993.
———. "Intentionality: Aquinas and Wittgenstein." In *Thomas Aquinas: Contemporary Philosophical Perspectives,* edited by Brian Davies, 243–56. New York: Oxford University Press, 2002.
Kerr, Fergus. *After Aquinas: Versions of Thomism.* Malden, Mass.: Blackwell, 2002.
Kim, Jaegwon. "Causality, Identity, and Supervenience in the Mind-Body Problem." *Midwest Studies in Philosophy* 4, no. 1 (1979): 31–49.
———. "On the Psycho-Physical Identity Theory." *American Philosophical Quarterly* 3 (1966): 227–35.
———. *The Philosophy of Mind.* Boulder, Colo.: Westview Publishing, 1998.
———. *The Philosophy of Mind.* 3rd ed. Boulder, Colo.: Westview Publishing, 2011.
———. *Physicalism or Something Near Enough.* Princeton, N.J.: Princeton University Press, 2005.
———. "Psychophysical Supervenience." *Philosophical Studies* 41 (1982): 51–70.
———. "Supervenience." In *A Companion to the Philosophy of Mind,* edited by Samuel Guttenplan, 575–83. Cambridge, Mass.: Blackwell, 1995.
———. *Supervenience and Mind.* Cambridge: Cambridge University Press, 1993.
King, Peter. "Medieval Intentionality and Psuedo-Intentionality." *Quaestio* 10 (2010): 25–44.
Kirkham, Richard L. *Theories of Truth: A Critical Introduction.* Cambridge, Mass.: MIT Press, 1992.
Klima, Gyula. "Aquinas on the Materiality of the Human Soul and the Immateriality of the Human Intellect." *Philosophical Investigations* 32, no. 2 (April 2009): 163–82.
Konyndyke, Kenneth. *Introductory Modal Logic.* Notre Dame, Ind.: University of Notre Dame Press, 1986.
Koons, Robert C. *Realism Regained: An Exact Theory of Causation, Teleology, and the Mind.* New York: Oxford University Press, 2000.
Kragh, Helge. *Cosmology and Controversy: The Historical Development of Two Theories of the Universe.* Princeton, N.J.: Princeton University Press, 1996.
Kreeft, Peter. *Socratic Logic.* Notre Dame, Ind.: St. Augustine's Press, 2005.
Kripke, Saul. *Naming and Necessity.* Cambridge, Mass.: Harvard University Press, 1980.

Kronen, John D., Sandra Mennsen, and Thomas D. Sullivan. "The Problem of the Continuant: Aquinas and Suarez on Prime Matter and Substantial Generation." *Review of Metaphysics* 53, no. 4 (2000): 863–85.

Latham, Noah. "What Is Token Physicalism?" *Pacific Philosophical Quarterly* 84, no. 3 (2003): 270–90.

Leftow, Brian. "Mind, Brain, Soul." In *The Waning of Materialism,* edited by Robert C. Koons and George Bealer, 395–417. New York: Oxford University Press, 2010.

———. "Souls Dipped in Dust." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons,* edited by Kevin Corcoran, 120–38. Ithaca, N.Y.: Cornell University Press, 2001.

Leibniz, G. W. "The Monadology." In *Philosophical Papers and Letters.* Translated by Leroy Loemker, 643–53. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1969.

Lewis, C. S. *The Abolition of Man.* In *The Complete C. S. Lewis Signature Classics,* 689–730. New York: Harper Collins, 2002.

———. *Miracles.* New York: Harper Collins, 1947.

Lewis, David. "Psychophysical and Theoretical Explanations." *Australasian Journal of Philosophy* 50, no. 3 (1972): 249–58.

———. "What Experience Teaches." In *Mind and Cognition,* edited by William Lycan, 499–518. Cambridge, Mass.: Blackwell, 1991.

Locke, John. *Essay Concerning Human Understanding.* Edited by P. Nidditch. Oxford: Claredon Press, 1975.

Loux, Michael. *Metaphysics: A Contemporary Introduction.* New York: Routledge, 1998.

———, ed. *The Possible and the Actual: Readings in the Metaphysics of Modality.* Ithaca, N.Y.: Cornell University Press, 1979.

Lowe, E. J. *Personal Agency: The Metaphysics of Mind and Action.* New York: Oxford University Press, 2008.

———. "Substance Dualism: A Non-Cartesian Approach." In *The Waning of Materialism,* edited by Robert Koons and George Bealer, 439–61. New York: Oxford University Press, 2010.

———. *A Survey of Metaphysics.* Oxford: Oxford University Press, 2002.

———. "Naturalism, Imagination, and the Scientific Worldview." In *Turning Images in Philosophy, Science, and Religion,* edited by Charles Taliaferro and Jil Evans, 91–100. New York: Oxford University Press, 2011.

Machuga, Ric. *In Defense of the Soul: What It Means to Be Human.* Grand Rapids, Mich.: Brazos Press, 2002.

MacIntyre, Alasdair. *Dependent Rational Animals: Why Human Beings Need the Virtues.* Chicago: Open Court, 1999.

Madden, James D. "The Fifth Way, Scientism, and Intelligent Design." *Faith and Reason* 31, no. 3 (Fall 2006): 387–408.

———. "Leibniz on Teleology and the Intelligibility of Nature." *Proceedings of the American Catholic Philosophical Association* 77 (2004): 173–88.
———. "Realism, Nominalism, and Biological Naturalism." *International Philosophical Quarterly* 51, no. 1 (2011): 85–102.
Madden, James D., and Mark A. Discher. "What Intelligent Design Does Not Imply." *Perspectives on Science and Christian Faith* 56 (December 2004): 286–81.
———. "What Would Count as Defeating Naturalism? A Reply to Van Till." *Perspectives on Science and Christian Faith* (December 2004): 296–98.
Malebranche, Nicolas. *The Search after Truth.* Translated by Thomas Lemmon. New York: Cambridge University Press, 1997.
Mankin, Stephen. "Aquinas, Natural Tendencies, and Natural Kinds." *New Scholasticism* 13, no.3 (1989): 253–74.
Maritain, Jacques. *The Degree of Knowledge.* Notre Dame, Ind.: University of Notre Dame Press, 1999.
———. *Introduction to Philosophy.* Translated by E. I. Watkins. New York: Rowman and Littlefield, 2005.
———. *The Philosophy of Nature.* New York: Philosophical Library, 1951.
McDowell, John. *Mind and World.* Cambridge, Mass.: Harvard University Press, 1996.
McGinn, Colin. "Anomalous Monism and Kripke's Cartesian Intuitions." *Analysis* 37, no. 2 (1977): 78–80.
———. "Can We Solve the Mind-Body Problem?" *Mind* 98, no. 391 (1989): 349–66. Reprinted in *Philosophy of Mind: Classical Problems/Contemporary Issues,* edited by Brian Beakely and Peter Ludlow, 321–36. Boston, Mass.: MIT Press, 2006.
Meehl, P. E., and Sellars, Wilfrid. "The Concept of Emergence." In *The Foundations of Science and the Concepts of Psychology and Psychoanalysis,* edited by Herbert Feigl and Michael Scriven, 239–52. Minnesota Studies in the Philosophy of Science 1. Minneapolis: University of Minnesota Press, 1956.
Melnyk, Andrew. "Can Physicalism Be Non-reductive?" *Philosophy Compass* 3, no. 6 (2008): 1281–96.
———. "Naturalism, Free Choices, and Conscious Experiences." http://www.infidels.org/library/modern/andrew_melnyk/against-dualism.html.
Menssen, Sandra, and Thomas Sullivan. *The Agnostic Inquirer: Revelation from a Philosophical Standpoint.* Grand Rapids, Mich.: Eerdmans Publishing, 2007.
Menuge, Angus. *Agents under Fire: Materialism and the Rationality of Science.* New York: Rowman and Littlefield, 2004.
Merricks, Trenton. *Objects and Persons.* New York: Oxford University Press, 2002.

———. "The Resurrection of the Body and Life Everlasting." In *Reason for the Hope Within,* edited by Michael Murray, 261–87. Grand Rapids, Mich.: Eerdmans Publishing, 1999.

Millikan, Ruth. "Biosemantics." In *The Philosophy of Mind: Classical and Contemporary Readings,* edited by David Chalmers, 500–508. New York: Oxford University Press, 2002.

———. "In Defense of Proper Functions." *Philosophy of Science* 56, no. 2 (1989): 288–302.

———. *Language, Thought and Other Biological Categories: New Foundations for Realism.* Cambridge, Mass.: MIT Press, 1984.

Minsky, Marvin. *The Society of Mind.* New York: Simon and Schuster, 1985.

Moody, Raymond. *Life after Life.* Atlanta, Ga.: Mockingbird Books, 1975.

Moreland, J. P. "Hume and the Argument from Consciousness." In *In Defense of Natural Theology: A Post-Humean Assessment,* edited by James Sennett and Doug Groothius, 271–96. Downers Grove, Ill.: InterVarsity Press, 2005.

———. *The Recalcitrant Imago Dei: Human Persons and the Failure of Naturalism.* London: SCM Press, 2009.

Murphy, Nancy. *Bodies and Souls, or Spirited Bodies.* Cambridge: Cambridge University Press, 2006.

Nagel, Thomas. "Armstrong on Mind." In *Readings in Philosophy of Psychology,* edited by Ned Block, 200–206. Cambridge, Mass.: Harvard University Press, 1980.

———. "Conceiving the Impossible in the Mind-Body Problem." *Philosophy* 73, no. 285 (1998): 337–52.

———. *The Last Word.* New York: Oxford University Press, 2001.

———. *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False.* New York: Oxford University Press, 2012.

———. *The View from Nowhere.* New York: Oxford University Press, 1986.

———. "What Is It Like to Be a Bat?" *Philosophical Review* 83, no. 4 (1974): 435–50. Reprinted in *The Philosophy of Mind: Classical and Contemporary Readings,* edited by David Chalmers, 219–25. Oxford: Oxford University Press, 2002.

Nemirow, Laurence. "Physicalism and the Cognitive Role of Acquaintance." In *Mind and Cognition,* edited by William Lycan, 490–99. Cambridge, Mass.: Blackwell, 1990.

Nietzsche, Friedrich. *The Gay Science.* Translated by Adrian Del Caro. Cambridge: Cambridge University Press, 2001.

O'Connor, Timothy. "Causality, Mind, and Free Will." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons,* edited by Kevin Corcoran, 44–58. Ithaca, N.Y.: Cornell University Press, 2001.

———. "Emergent Properties." *American Philosophical Quarterly* 31, no. 2 (April 1994): 91–104.
———. *Persons and Causes.* New York: Oxford University Press, 2000.
———. "Review of *All the Power in the World* by Peter Unger." *Notre Dame Philosophical Reviews,* March 10, 2007. http://ndpr.nd.edu/news/25237-all-the-power-in-the-world/.
———. *Theism and Ultimate Explanation: The Necessary Shape of Contingency.* Malden, Mass.: Blackwell, 2008.
Oderberg, David, ed. *Form and Matter.* Oxford: Blackwell, 1999.
———. *Real Essentialism.* New York: Routledge, 2007.
Olsen, Eric. *What Are We? A Study in Personal Ontology*. New York: Oxford University Press, 2007.
Owens, Joseph. "Common Nature: A Point of Comparison between Thomistic and Scotistic Metaphysics." *Mediaeval Studies* 19 (1957): 1–40.
Pannier, Russell, and Thomas D. Sullivan, "Consciousness and Intentional Awareness of Instantiables." In *Mind and Its Place in the World: Non-reductionist Approaches in the Ontology of Consciousness,* edited by Alexander Batthyany and Avshalom Elitzur, 77–100. Frankfurt: Ontos Verlag, 2006.
Pasnau, Robert. *Thomas Aquinas on Human Nature: A Philosophical Study of Summa Theologiae 1a 75–89.* New York: Cambridge University Press, 2002.
Peifer, John Fredrick. *The Concept in Thomism.* New York: Bookman, 1952.
Place, U. T. "Is Consciousness a Brain Process?" *British Journal of Psychology* 47 (1956): 44–50. Reprinted in *The Philosophy of Mind: Classical and Contemporary Readings,* edited by David Chalmers, 55–60. New York: Oxford University Press, 2002.
Plantinga, Alvin. *Warranted Christian Belief.* New York: Oxford University Press, 2000.
———. *Where the Conflict Really Lies: Science, Religion, and Naturalism.* New York: Oxford University Press, 2011.
Plantinga, Alvin, and Daniel Dennett. *Science and Religion: Are They Incompatible?* New York: Oxford University Press, 2011.
Plantinga, Alvin, and Michael Tooley. *Knowledge of God.* Malden, Mass.: Blackwell, 2008.
Plato. *Phaedo.* Translated by G. M. A. Grube. Indianapolis, Ind.: Hackett Publishing, 1978.
Polger, Thomas. "Identity Theories." *Philosophy Compass* 4, no. 5 (2009): 822–34
Popper, Karl, and John Eccles. *The Self and Its Brain: An Argument for Interactionism.* New York: Routledge, 1977.
Price, H. H. "Universals and Family Resemblances." In *Classics in Analytical*

Metaphysics, edited by Larry Lee Blackman, 211–32. Lanham, Md.: University Press of America, 1984.
Post, John. *Metaphysics: A Contemporary Introduction.* New York: Paragon, 1991.
Pouivet, Roger. *After Wittgenstein, St. Thomas.* South Bend, Ind.: St. Augustine Press, 2006.
Pruss, Alexander. "What Are Aristotelian Forms?" http://bearspace.baylor.edu/Alexander_Pruss/www/papers/Forms.html.
Putnam, Hillary. "Brains and Behavior." In *Philosophy of Mind: Classical and Contemporary Readings,* edited by David Chalmers, 45–54. New York: Oxford University Press, 2002.
———. "Is Water Necessarily H_2O?" In *Realism with a Human Face,* edited by James Conat, 54–79. Cambridge, Mass.: Harvard University Press, 1990.
———. "The Meaning of 'Meaning.'" In *Language, Mind, and Knowledge,* edited by K. Gunderson, 131–93. Minnesota Studies in the Philosophy of Science 7. Minneapolis: University of Minnesota Press, 1975.
———. "Psychological Predicates." *In Art, Mind, and Nature,* edited by W. Captain and D. Merrill, 229–440. Pittsburgh, Pa.: University of Pittsburgh Press, 1967.
———. *Renewing Philosophy.* Cambridge, Mass.: Harvard University Press, 1995.
———. *The Threefold Cord: Mind, Body, and World.* New York: Columbia University Press, 1999.
———. *Words and Life.* Cambridge, Mass.: Harvard University Press, 1994.
Quine, W. V. O. "Epistemology Naturalized." In *Ontological Relativity and Other Essays,* 69–91. New York: Columbia University Press, 1969.
Rea, Michael. *Material Constitution: A Reader.* New York: Rowman and Littlefield, 1996.
———. *A World without Design: The Ontological Consequences of Naturalism.* New York: Oxford University Press, 2002.
Reichmann, James B. *Evolution, Animal "Rights," and the Environment.* Washington, D.C.: The Catholic University of America Press, 2000.
Reitan. Eric. *Is God a Delusion? A Reply to Religion's Cultured Despisers.* Malden, Mass.: Wiley-Blackwell, 2009.
Reppert, Victor. *C. S. Lewis's Dangerous Idea: In Defense of the Argument from Reason.* Downers Grove, Ill.: InterVarsity Press, 2003.
Robinson, Daniel N. *Consciousness and Mental Life.* New York: Columbia University Press, 2008.
Rorty, Richard. *Contingency, Irony, and Solidarity.* Cambridge: Cambridge University Press, 1989.
———. "Mind-Body Identity, Privacy, and Categories." *Review of Metaphysics*

19, no. 1 (September 1965): 24–54. Reprinted in *Materialism and the Mind-Body Problem*, edited by David M. Rosenthal, 174–99. Indianapolis, Ind.: Hackett Publishing, 2000.

———. *Philosophy and the Mirror of Nature.* Princeton, N.J.: Princeton University Press, 1979.

Rosenblum, Bruce, and Fred Kuttner. *Quantum Enigma: Physics Encounters Consciousness.* New York: Oxford University Press, 2006.

Rosenthal, David M., ed. *Materialism and the Mind-Body Problem.* Indianapolis, Ind.: Hackett Publishing, 2000.

Ross, James. *Thought and World: The Hidden Necessities.* Notre Dame, Ind.: University of Notre Dame Press, 2008.

Rowe, William L. *The Cosmological Argument.* New York: Fordham University Press, 1998.

Rudder Baker, Lynn. *Persons and Bodies: A Constitution View.* Cambridge: Cambridge University Press, 2000.

Ruse, Michael. "Evolutionary Theory and Christian Ethics." In *The Darwinian Paradigm,* 251–72. London: Routledge, 1989.

Russell, Bertrand. *Portraits from Memory.* Nottingham: Spokesman, 1956.

———. *What I Believe.* New York: Routledge, 2004.

Ryle, Gilbert. *The Concept of Mind.* Chicago: University of Chicago Press, 1949.

Sabom, Michael. *Recollections of Death: A Medical Investigation.* New York: Harper and Row, 1982.

Sachs, Oliver. *The Man Who Mistook His Wife for a Hat: And Other Clinical Tales.* New York: Touchstone, 1998.

Schornborn, Christopher Cardinal. *Chance or Purpose: Creation, Evolution, and a Rational Faith.* San Francisco: Ignatius Press, 2007.

Searle, John. *Freedom and Neurobiology: Reflections on Free Will, Language, and Political Power.* New York: Columbia University Press, 2007.

———. *Mind: A Brief Introduction.* New York: Oxford University Press, 2004.

———. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 1 (1980): 417–24.

———. *Minds, Brains, and Science.* Cambridge, Mass.: Harvard University Press, 1984.

———. *The Mystery of Consciousness.* New York: New York Review Books, 1997.

———. *Philosophy in a New Century: Selected Essays.* Cambridge: Cambridge University Press, 2008.

———. *Rationality in Action.* Boston, Mass.: MIT Press, 2003.

———. *The Rediscovery of the Mind.* Cambridge, Mass.: MIT Press, 1992.

Sellars, Roy Wood. *Evolutionary Naturalism.* Chicago: Open Court Press, 1922.

Sellars, Wilfrid. "Being and Being Known." In *Science, Perception, and Reality,*

41–59. Atascadero, Calif.: Ridgeview Publishing, 1991.
———. "Empiricism and the Philosophy of Mind." In *Science, Perception, and Reality,* 127–96. Atascadero, Calif.: Ridgeview Publishing, 1991.
———. "Philosophy and the Scientific Image of Man." In *Science, Perception, and Reality,* 1–40. Atascadero, Calif.: Ridgeview Publishing, 1991.
Sider, Ted. "What's So Bad about Overdetermination?" *Philosophy and Phenomenological Research* 67, no. 3 (2003): 719–26.
Simon, Yves. *The Great Dialogue of Nature and Space.* Notre Dame, Ind.: St. Augustine's Press, 1970.
Simons, Peter. "Farewell to Substance: A Differentiated Leave-Taking." In *Form and Matter,* edited by David S. Oderberg, 22–39. Malden, Mass.: Blackwell, 1999.
Skinner, B. F. *Beyond Freedom and Dignity.* Indianapolis, Ind.: Hackett Publishing, 2002.
Smart, J. J. C. "Sensations and Brain Processes." *Philosophical Review* 68 (1959): 141–56. Reprinted in *Materialism and the Mind-Body Problem,* 2nd ed., edited by David M. Rosenthal, 53–66. Indianapolis, Ind.: Hackett Publishing, 2000.
Soames, Scott. *Philosophical Analysis in the Twentieth-Century.* Vol. 2. Princeton, N.J.: Princeton University Press, 2003.
———. *Reference and Description.* Princeton, N.J.: Princeton University Press, 2005.
Sober, Elliot. "Putting the Function Back into Functionalism. In *Mind and Cognition,* edited by William G. Lycan, 97–106. Cambridge, Mass.: Blackwell, 1990.
Sperry, Roger. "In Defense of Mentalism and Emergent Interaction." *Journal of Mind and Behavior* 12, no. 2 (Spring 1991): 221–45.
Stace, Walter. "Naturalism and Religion." *Proceedings and Addresses of the American Philosophical Association* 23 (1949–50): 22–46.
Steinihard, Eric. *More Precisely: The Math You Need to Do Philosophy.* Peterboro, Ontario: Broadview Press, 2009.
Stork, Michael. "Parts, Wholes, and Presence by Power: A Response to Gordon P. Barnes." *Review of Metaphysics* 62 (2008): 45–59.
Strawson, Galen. *Real Materialism and Other Essays.* New York: Oxford University Press, 2008.
Strawson, Peter. *Individuals: An Essay in Descriptive Metaphysics.* New York: Routledge, 1964.
Stump, Eleonore. *Aquinas.* London: Routledge, 2003.
———. "Non-Cartesian Substance Dualism and Materialism without Reductionism." *Faith and Philosophy* 12, no. 4 (October 1995): 505–31.
———. "Resurrection, Reassembly, and Reconstitution: Aquinas on the Soul." In *Die menschliche Seele: Brauchen wir den Dualismus?* edited by

Bruno Niederberger and Edmund Runggaldier, 151–72. Frankfort: Ontos Verlag, 2006.

Suarez, Francisco. *On Creation, Conservation, and Concurrence: Metaphysical Disputations 20–22.* Translated by A. J. Freddoso. South Bend, Ind.: St. Augustine's Press, 2002.

Swinburne, Richard. *The Evolution of the Soul.* Rev. ed. New York: Oxford, 1997.

———. *Is There a God?* New York: Oxford University Press, 1997.

Taliaferro, Charles. *Consciousness and the Mind of God.* Cambridge: Cambridge University Press, 1994.

———. "Possibilities in the Philosophy of Mind." *Philosophy and Phenomenological Research* 5, no. 1 (1997): 127–37.

———. "The Virtues of Embodiment." *Philosophy* 76, no. 295 (January 2001): 111–25.

Taliaferro, Charles, and Jil Evans. *The Image in Mind: Theism, Naturalism, and the Imagination.* London: Continuum Publishing, 2011.

Toner, Patrick. "Emergent Substance." *Philosophical Studies* 141, no. 3 (2008): 281–97.

———. "Hylomorphic Animalism." *Philosophical Studies* 155 (2011): 65–81.

———. "St. Thomas on Death and the Separated Soul." *Pacific Philosophical Quarterly* 91 (2010a): 587–99.

———. "Thomas Versus Tibbles: A Critical Study of Brown's Aquinas and the Ship of Theseus." *American Catholic Philosophical Quarterly* 81, no. 4 (2007): 639–53.

Tooley, Michael, Celia Wolf-Devine, Philip Devine, and Alison M. Jaggar. *Abortion: Three Perspectives.* Oxford: Oxford University Press, 2009.

Turing, Alan. "Computer Machinery and Intelligence." In *Mind Design 2,* edited by J. Haugeland, 29–56. Cambridge, Mass.: MIT Press, 1997.

Tye, Michael. "Naturalism and the Problem of Intentionality." *Midwest Studies in Philosophy* 19 (September 1994): 122–42.

Unger, Peter. *All the Power in the World.* New York: Oxford University Press, 2006.

———. "The Problem of the Many." *Midwest Studies in Philosophy* 5 (1980): 411–67.

Van Inwagen, Peter. "I Look to the Resurrection of the Dead and Life in the World to Come." http://philosophy.nd.edu/people/all/profiles/van-inwagen-peter/documents/Resurrection.doc.

———. *Material Beings.* Ithaca, N.Y.: Cornell University Press, 1990.

———. *Metaphysics.* Boulder, Colo.: Westview Publishing, 2009.

———. "Ontological Arguments." *Noûs,* 11, no. 4 (1977): 375–95.

Varghese, Roy Abraham, ed. *Great Thinkers on Great Questions.* Oxford: One World Publishing, 1998,

Wallace, William A. *The Modeling of Nature: Philosophy of Science and Philosophy of Nature in Synthesis.* Washington, D.C.: The Catholic University of America Press, 1996.

Weatherson, Brian. "The Problem of the Many." In *The Standford Encyclopedia of Philosophy.* http://plato.stanford.edu/entries/problem-of-many/.

Wiker, Benjamin. *Moral Darwinism: How We Became Moral Hedonists.* Downers Grove, Ill.: InterVarsity Press, 2002.

Wigner, Eugene. "Remarks on the Mind-Body Question." In *Philosophy of Mind: A Guide and an Anthology,* edited by John Heil, 866–77. Oxford: Oxford University Press, 2004.

Yandell, Keith. "A Defense of Dualism." *Faith and Philosophy* 12, no. 4 (October 1995): 548–66.

Zimmerman, Dean. "Christians Should Affirm Mind-Body Dualism." In *Contemporary Debates in Philosophy of Religion,* edited by Michael L. Peterson and Raymond J. Vanarragon, 314–27. Malden, Mass.: Routledge, 2004.

———. "From Experience to Experience." In *The Soul Hypothesis: Investigations into the Existence of the Soul,* edited by Mark C. Baker and Stewart Goetz, 168–201. New York: Continuum, 2011.

INDEX

accidental change, 223–25; atomism and, 230–31; hylomorphism and, 236–37
accidental property, 223
Anscombe, Elizabeth, 248n8; on unintelligibility of immaterial substance, 25–27
Aquinas, St. Thomas: as emergent property dualist, 262–64; on intentionality as formal identity, 260–64; on irreducibility of thought, 203–9; on moral agency, 278–80; on the problem of change, 235–42; rejection of substance dualism, 270–71; on sensation, 255–64; on separability/immortality of the human soul, 265–71; on universals, 207–8; on virtual presence of the elements, 241–43, 247, 258–59, 275, 279, 284. *See also* hylomorphism
Aristotle: on causality, 243–49; on the problem of change, 235–42; on the soul as the form of a living substance, 251–55. *See also* hylomorphism
artificial intelligence, 128–31; Chinese Room argument against, 147–52
atomism, 222, 229–35

behaviorism, 95–103; and functionalism, 125. *See also* materialism
biological naturalism, 170–74, 178–82, 193n39, 207n62, 212, 216. *See also* Searle, John
Bobik, Joseph, 222n3, 259n13, 259n39
Broad, C. D., 190n3
Brower, Jeffrey, 237–39, 282

Cartesian dualism, ix–x, 30, 83. *See also* Descartes, Rene; substance dualism
causation, 196, 221n2, 243–49, 264; mental, 13–15, 60–68, 101, 125–26, 275
Chalmers, David, 38n32, 42n36; on evolution and qualia, 198–201; on panpsychism, 189n29; on qualia, 147; on zombies, 137–38, 200–204
change, problem of, 222–42. *See also* accidental change; substantial change
Churchland, Patricia, 91n4
Churchland, Paul, 90–94
Chinese Room argument, 147–63, 213
Clarke, W. Norris, 8n18, 221n2
conceivability, 41–47, 106, 196. *See also* modal argument
conservation laws, 65–69, 275

Darwinian evolution: and agency, 165–67, 210–16; and the human soul, 272–74; and intentionality, 155–158; and qualia, 198–201
Davidson, Donald, 118–21
Dennett, Daniel, 8n17, 16, 126n60, 143–47, 158–161, 215n75
Descartes, Rene, ix, 22n7, 27, 30–31, 35, 38n29, 41, 46–47, 51n55, 61–63, 84, 106–7, 173, 221, 276
disposition, 98
dualism, 9–20; contrasted with hylomorphism, 274–78; difference (simplicity/unity) argument for, 47–56; emergent dualism, 20n1, 83n55, 171n1, 186n25, 191, 272n36; modal argument for, 30–46; objections to, 25–30, 60–88; plausibility of substance dualism, 217–21; property dualism, 9–20, 38–40, 61, 81–82, 109–13, 120–21, 138n10, 141, 180, 193n39

eliminative materialism, 36n26, 89–95, 110, 132, 160, 169, 132. *See also* materialism
emergentism, 171–220; and dualism, 20n1, 272; hylomorphism as a version of,

emergentism *(cont.)*
256–64; predictive or nomological emergence of the human soul, 274
Epicurus, 192n37, 229
epiphenomenalism, 60–61, 72–73, 101, 176, 211
essential property, 223–24

Feser, Edward, 137n8, 156n37–38, 161, 164, 280n46
folk psychology, 90–95, 119
form, 235–36; accidental, 236; ambiguity between abstract and concrete, 280–84; as cause, 243; as co-principle, 239; defined functionally, 237–38, 282–83; as principle of change, 235–36; as soul of a living thing, 251–55; subsistent, 239, 270, 275, 280–84; substantial, 236
Freddoso, A. J., 64–65, 242n28, 248n38
functionalism, 122–33, 147, 170, 216; contrasted with materialism, 127–28; and intentionality, 147–58; and rational agency, 161–64

Geach, Peter, 43–7, 99n19, 232–33n17, 271n35
George, Robery, 84–85, 204n57, 237n19, 256, 266n25, 270–72, 278n42
Gilson, Etienne, 222n3, 258, 274n39

Haldane, John, 12, 28–29, 154n34, 203n57, 206, 207–9, 240n24, 264–68, 271n34; on identity of human person with body, 83–59, 270–71, 276
Hume, David, 2, 39, 64–65, 171n1, 195–97, 247–49

identity, 27–28, 34–37, 47–50, 89; contingent (empirical), 103–6; token, 113–22, 132–33, 174n7, 175; type, 114–19, 123, 132–33, 174–75
identity theory, 102–13; and eliminative materialism, 113; and neural chauvinism, 115–18; and property dualism, 109–13; replies to dualist arguments, 106–8; and topic neutrality, 110–13. *See also* identity; materialism
intentionality, 12–15, 147–61; and emergentism, 203–9; evolutionary accounts, 155–58; and hylomorphism, 255–64; instrumental account, 158–61; and materialism, 147–61

Jackson, Frank, 42n36, 135–40

Kant, Immanuel, 55n51
Kenny, Anthony, 262–64, 266n26, 270n32, 280n45, 281, 283n52
Kim, Jaegwon, 60–61, 69–73, 109, 118–23, 126n60, 127n62, 143
knowledge argument, 30n20, 133–43. *See also* qualia
Kripke, Saul, 35n25, 105, 117n43

Lee, Patrick, 204n57, 237n19, 256, 266n25, 270–72, 278n42
Leftow, Brian, 259n12, 272n37
Leibniz. G. W., 48–51, 62
Lewis, C. S., 2–3, 6, 17n23, 164–66, 202n55
Lewis, David, 123n57, 126n60, 142
Locke, John, 171n1
Lowe, E. J., 8n18, 10n20, 28n17, 51n56, 62n9, 67n21, 77–80, 160n43, 165n47, 171n1, 180n18, 221n2, 232–34, 240n24, 242n29

MacIntyre Alasdair, 85n60, 273n38
Maritain, Jacques, 221n2
materialism, ix–x, 5–6, 88–131; contrasted with functionalism, 127–28; nonreductive version, 114–31, 170; objections to, 132–68. *See also* behaviorism; eliminative materialism; identity theory
matter: Cartesian conception of, 173–74; as co-principle, 239; identity and persistence, 27–29; as principle of change, 235–36. *See also* prime matter
McGinn, Colin, 40n32, 117n43, 194, 202
mechanism, xi, 21–22, 27, 166, 172–73, 191, 219–21, 258, 277; contrasted with hylomorphic philosophy of nature, 243–50, 260; and property dualism, 39–40; as shared assumption of dualists and materialists, 21–23; as source of difficulty for emergentism, 197–8, 257–58; and unintelligibility of causality, 64–65

Menssen, Sandra, 27n16, 64–65, 67n20, 68–69
mysterianism, 194–95

Nagel, Thomas, 40n32, 42n36, 140, 189n29, 193n39; on naturalism and religion, 202; on qualia, 133–38
naturalism, 1–7; as unquestioned assumption, 201–3
neural chauvinism. *See* identity theory

occasionalism, 62
Oderberg, David, 39n30, 154n34, 160n43, 223n4, 232n14, 257n10

panpsychism, 56n65, 188–90; contrasted with hylomorphic account of matter, 260
Pasnau, Robert, 208–9n65, 239n23
physical property, 10
Plantinga, Alvin, 7–8, 40n33, 50, 69n27, 80n47, 156–57, 165n47, 171n1, 193n39, 215–16
Plato, 171n1, 205n59, 221n3, 284
Popper, Karl, 6n14, 62n8, 67n20, 69, 171n1, 246n36
prime matter, 237–48, 251–52, 259–60, 275, 284
Putnam, Hilary, 101, 105n26, 115n41, 123n57, 157–58, 207–8n62, 246, 265n21

qualia, 30n20, 50n51, 61n3, 155, 160, 195, 218; and arguments against materialism, 133–38; evolutionary role, 200; materialist treatments, 138–47. *See also* knowledge argument; sensation
quantum physics, 28, 62–68, 211

reduction, 103n47, 118–19, 138; Searle on causal vs. ontological, 174–75. *See also* identity theory
resurrection, 271
rigid designator, 105
Russell, Bertrand, 6, 27
Ryle, Gilbert, 95n15

Searle, John, 40n32, 61n5, 66, 82, 92, 107–9, 116–17, 129, 140–41, 167–8, 256, 259n13; on behaviorism, 99–102; biological explanation for emergence of consciousness, 199; on biological naturalism as neither materialism nor dualism, 178–82; on causal theories of intentionality, 154–56; on Chinese Room argument, 150–2; criticism of his emergentism, 182–96; on naturalism and free agency, 210–12; on panpsychism, 189; on philosophy of mind as based on false assumptions, 170–78; rejection of supernaturalism, 201. *See also* biological naturalism
Sellars, Wilfrid, 5, 43n38, 47n45, 90n2, 92n6, 95n15, 171n1, 191n15, 207n62, 265n21
Simon, Yves, 232
Skinner, B. F., 17n28, 95
soul: nonevolutionary origin of the human, 272–74; not identical to human person, 271–72; subsistence/immortality of human, 203–9, 265–70; as substantial form of a living substance, 251–55
Strawson, Galen, 27n16, 40n32, 146, 173, 187, 191–93, 196, 257
Strawson, Peter, 40n33
substantial change, 223–24; atomism and, 230–32; hylomorphism and, 236–38
Sullivan, Thomas, 204–5, 239n23
supervenience, 121–22, 132–33, 260–61, 270; distinguished from emergentism, 171n1, 174, 179, 182–83, 220; and functionalism, 122–27; and normativity, 166–67; and qualia, 133–38
Swinburne, Richard, 86n61, 171n1, 180n18

Toner, Patrick, xii, 40n32, 200n50, 242n30, 266n25, 271n34
topic neutrality, 110–13. *See also* identity theory
Turing, Alan, 130
Turing test, 130, 151, 164

virtual presence 241–43, 247, 258–59, 275, 279, 284

zombies, 30n20, 136–37, 142, 200